

2005

Genetic analysis and molecular characterization of  
the W4 locus and the k2 Mdh1-n y20  
chromosomal region in soybean [Glycine max (L)  
Merr]

Min Xu

*Iowa State University*

Follow this and additional works at: <https://lib.dr.iastate.edu/rtd>



Part of the [Genetics Commons](#)

### Recommended Citation

Xu, Min, "Genetic analysis and molecular characterization of the W4 locus and the k2 Mdh1-n y20 chromosomal region in soybean [Glycine max (L) Merr] " (2005). *Retrospective Theses and Dissertations*. 1607.  
<https://lib.dr.iastate.edu/rtd/1607>

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact [digirep@iastate.edu](mailto:digirep@iastate.edu).

Genetic analysis and molecular characterization of the *W4* locus and the *k2 Mdh1-n y20*  
chromosomal region in soybean [*Glycine max* (L.) Merr.]

by

Min Xu

A dissertation submitted to the graduate faculty  
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Genetics

Program of Study Committee:  
Reid G. Palmer, Major Professor  
Madan K. Bhattacharyya  
E. Charles Brummer  
David Grant  
Thomas A. Peterson

Iowa State University

Ames, Iowa

2005

Copyright © Min Xu, 2005. All rights reserved.

UMI Number: 3184665

### INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

**UMI<sup>®</sup>**

---

UMI Microform 3184665

Copyright 2005 by ProQuest Information and Learning Company.

All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company  
300 North Zeeb Road  
P.O. Box 1346  
Ann Arbor, MI 48106-1346

Graduate College  
Iowa State University

This is to certify the dissertation of

Min Xu

has met the dissertation requirements of Iowa State University

Signature was redacted for privacy.

Major Professor

Signature was redacted for privacy.

For Major Program

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS	v
ABSTRACT	vi
CHAPTER 1	GENERAL INTRODUCTION
	1
1. Introduction	1
2. Dissertation organization	4
3. Literature review	5
3.1. Transposable elements in plant genomes	5
3.1.1. RNA-mediated transposable elements	5
3.1.2. DNA-mediated transposons	9
3.2. Transposable elements as mutagens	11
3.2.1. Small-scale mutations	11
3.2.2. Large-scale mutations	12
3.3. Transposable elements in functional genomics	13
3.4. Transposable element studies in soybean	15
4. References	18
CHAPTER 2	GENETIC ANALYSIS AND MOLECULAR MAPPING OF A PALE FLOWER ALLELE AT THE <i>W4</i> LOCUS IN SOYBEAN
	28
Abstract	28
Introduction	29
Materials and methods	31
Results	34
Discussion	38
Acknowledgements	41
References	41
CHAPTER 3	THE <i>W4</i> LOCUS CORRESPONDS TO A DIHYDROFLAVONOL-4-REDUCTASE (DFR) GENE IN SOYBEAN [ <i>GLYCINE MAX</i> (L.) MERR.]
	51
Abstract	51
Introduction	52
Materials and methods	53
Results	57
Discussion	61
Acknowledgements	67
References	67

CHAPTER 4	GENETIC ANALYSIS OF FOUR NEW MUTANTS AT THE UNSTABLE <i>K2 MDH1-N Y20</i> CHROMOSOMAL REGION IN SOYBEAN [ <i>GLYCINE MAX</i> (L.) MERR.]	78
	Abstract	78
	Introduction	79
	Materials and methods	81
	Results	83
	Discussion	85
	Acknowledgement	88
	References	88
CHAPTER 5	MOLECULAR MAPPING OF <i>K2 MDH1-N Y20</i> , AN UNSTABLE CHROMOSOMAL REGION IN SOYBEAN [ <i>GLYCINE MAX</i> (L.) MERR.]	99
	Abstract	99
	Introduction	100
	Materials and methods	101
	Results	104
	Discussion	108
	Acknowledgements	113
	References	114
CHAPTER 6	A LARGE DELETION IS INVOLVED IN THE <i>MDH1-N Y20</i> MUTATION IN SOYBEAN [ <i>GLYCINE MAX</i> (L.) MERR.]	127
	Abstract	127
	Introduction	128
	Materials and methods	130
	Results and discussion	132
	Conclusions	138
	Acknowledgements	139
	References	139
CHAPTER 7	GENERAL CONCLUSION	151

## **ACKNOWLEDGEMENTS**

I thank my major professor Dr. Reid G. Palmer for instructions and guidances during my Ph D process. I also like to thank all my present and former POS committee members: Dr. Madan K. Bhattacharyya, Dr. E. Charles Brummer, Dr. David Grant, Dr. Thomas A. Peterson, Dr. Steven S. Rodermel, and Dr. Daniel F. Voytas for providing instructions in this research. In addition, I would like to thank Dr. M. Paul Scott and Dr. Sandhu Devinder for helping in many experiments.

Finally, I would like to thank my present and former colleagues (Tracy Dang, Innan M. Cervantes, Liying Fan, and Pinar Uncu Brummer), my friends, and my family for their help and support, especially my husband Aigen Fu and my friend Dr. Huixia Shou who gave me a lot of advice and help in my project.

## ABSTRACT

The *W4* locus is an important locus to the anthocyanin biosynthesis in soybean flowers and hypocotyls. Mutations at this locus would cause less-pigmented flowers and hypocotyls, such as near-white flowers conditioned by the *w4* allele. Its mutable allele (*w4-m*) was suspected to contain an autonomous transposon. In this study, we reported a mutation with pale flowers and green hypocotyls conditioned by a new allele at the *W4* locus, *w4-p*. The dominance of these four alleles is: *W4* > *w4-m* > *w4-p* > *w4*. The inbred mutant line (*w4-p w4-p*) was assigned Genetic Type Collection number T369. The *W4* locus was positioned on the MLG D2 between Satt386 and nearby telomere with 2.3 cM genetic distance from Satt386.

The major components of anthocyanins in soybean flowers were found to be delphinidin and its derivatives, which were reduced in less pigmented flower petals caused by mutations at the *W4* locus. Results in this study showed that the pigment reduction was due to low transcript level of the DFR 2 (dihydroflavonol reductase 2) gene. The results also indicated that the *W4* gene very likely encoded the DFR2 protein.

Three closely linked genes *k2*, *Mdh1-n*, and *y20* form an unstable chromosomal region. The *Mdh1-n* mutation in T261 and all the *Mdh1-n y20* mutations were suggested to be due to genomic deletions. Four more new mutants in this region were reported in this study. One had tan-saddle seed coat (*k2*) isolated from a bulk harvest of cv. Kenwood. Three independent mutants (*k2 Mdh1-n y20*) were reported in the instability experiments. The unstable region was located on MLG H near SSR markers Satt253, Satt279, and Satt314. In addition, an extra putative deletion that covered Satt253, Satt279, and Satt314 was identified in T261. Mapping results indicated the two deletions were separate.



An *Mdh1* contig with size around 96 kb was constructed. Three copies of truncated *Calypso5-1* like retroelements were found downstream of the *Mdh1* gene (AF180335), and clustered in a ~16 kb chromosomal region. The deletions in all the *Mdh1*-n or *Mdh1*-n y20 mutants examined were larger than ~70 kb, but the breakpoints couldn't be identified.

## CHAPTER 1 GENERAL INTRODUCTION

### 1. Introduction

Several regions that were found in the soybean genome were very unstable, for example, the mutable allele of the *W4* locus (*w4-m*) and the *k2 Mdh1-n y20* chromosomal region. What causes the instability is unknown.

The *W4* locus conditions purple flowers and hypocotyls under *W1*\_ genetic background. Its recessive mutation *w4* gives near-white flowers and green hypocotyls (Groose and Palmer 1991). In 1983, a mutant with variegated flowers and hypocotyls was identified that was controlled by a mutable allele at the *W4* locus (*w4-m*). The mutable allele was dominant to the *w4* allele but recessive to the wild-type *W4* allele (Palmer et al. 1989). It could revert to the dominant *W4* allele somatically and germinally (Groose et al. 1988, 1990). The variegated phenotypes were the result of somatic reversion. Germinal reversion of the *w4-m* allele would cause plants to produce fully purple flowers and hypocotyls similar to wild type phenotypes. More interestingly, many new mutations were identified among the descendants of these wild-type germinal revertants such as chlorophyll-deficient leaves, necrotic roots, and complete or partially sterile plants (Palmer et al. 1989). Occasionally, germinal revertants would produce intermediate colored flowers such as dilute purple and pale (Groose et al. 1990).

The mutants, found in independent germinal revertants of the *w4*-mutable line, were mapped at different positions in the soybean genome. For example, mutant female partial-sterile 1 (*Fsp2*), female partial-sterile 2 (*Fsp3*), female partial-sterile 3 (*Fsp4*), and female partial-sterile 4 (*Fsp5*) were located on MLG C2, A2, F and G, respectively (Kato and

Palmer 2004). Mutant male-sterile and female-sterile (*st8*) was positioned on MLG J (Kato and Palmer 2003).

The instability of the *w4-m* allele was suspected to result from an active transposon residing adjacent to or within the *W4* locus (Palmer et al. 1989). The phenotypic reversion was the result of excision of the transposable element from this locus. The genetic stock for the true breeding mutant line (*w4-mw4-m*) was registered as the *w4*-mutable line, and was assigned Genetic Type Collection number T322 (Palmer et al. 1990).

Two mutant lines with intermediate colored flowers (dilute purple and pale) had been identified from the *w4*-mutable lines. The dilute-purple-mutant line was conditioned by a new allele at the *W4* locus (*w4-dp*), and was assigned Genetic Type Collection number T321 (Palmer and Goose 1993). It was suspected that the *w4-dp* mutation resulted from imprecise excision of the transposon or reinsertion into the regulatory region of the *W4* locus of the excised transposon.

Another unstable region consisted of three closely linked genes *k2*, *Mdh1-n*, and *y20* (Chen and Palmer 1998b). The *k2* locus conditions tan-saddle seed coat, the *Mdh1-n* locus conditions mitochondrial malate dehydrogenase (MDH, EC 1.1.1.37) 1 null, and the *y20* locus conditions chlorophyll-deficient foliage. The *y20* locus always co-segregates with the *Mdh1-n* locus (Palmer 1984; Chen and Palmer 1996, 1998a). The recombination rate between *k2* and *Mdh1-n y20* normally varies from 0% to 3% (Chen and Palmer 1996, 1998a). Southern blot results indicated that the co-segregation between *Mdh1-n* and *y20* loci was due to chromosome deletion (Imsande et al. 2001).

So far, 31 mutants were reported in this region (Palmer et al. 2004). Among these mutants, T323, T324, T325, and T346 (*Mdh1-n y20*) were found in the self-pollinated

progenies of germinal revertants of the *w4*-mutable line (T322), which was proposed to contain an autonomous transposable element at the *W4* locus (Palmer et al. 1989; Chen et al. 1999). T317 and T361 (*Mdh1-ny20*) were found among tissue culture-derived plants (Amberger et al. 1992; Palmer et al. 2000). Twelve mutants with a triple mutation (*k2 Mdh1-n y20*) were identified in the F<sub>2</sub> descendants derived from crosses of T239 (*k2*) or T261 (*k2 Mdh1-n*) with the wild-type parental strains of the *w4-m* and *Y18-m* mutable lines (Chen and Palmer 1998b).

The hypothesis for the instability of the *k2 Mdh1-n y20* chromosomal region is that a non-autonomous transposable element could reside adjacent to or in this region (Chen and Palmer 1998b). The non-autonomous transposable element could excise from its chromosomal position and cause mutations such as chromosome deletions, when it is activated by tissue culture or by transposases provided *in trans* through transposon tagging experiments or crossing experiments. Recently, it was found that in *Arabidopsis*, excision of a hybrid *Dissociation* (*Ds*) element could induce a genomic deletion flanking *Ds* with the size as large as 104 kb (Page et al. 2004).

In the present study, the objectives were to: 1) position the *W4* locus and the *k2 Mdh1-n y20* region on the soybean integrated genetic linkage map; 2) identify the *W4* gene through biochemical and molecular biological methods for further cloning the *W4* gene and the putative transposable element; 3) generate more mutants at the *k2 Mdh1-n y20* chromosomal region; and 4) identify the breakpoints of deletions in the *k2 Mdh1-n y20* mutants to determine what is responsible for the instability of this chromosome region.

## 2. Dissertation organization

This dissertation is presented in an alternative format and consists of seven chapters. Chapter 1 is a general introduction, in which, a literature review related to transposon studies with references cited is presented. Chapters 2 and 3 present the studies on the *W4* locus in soybean. Chapter 2 is a paper about positioning the *W4* locus on the soybean integrated map with simple sequence repeat (SSR) markers. This paper was published in the April 2005 issue of *Genome*. Chapter 3 presents the studies on identifying the function of the *W4* locus with biochemical and molecular biological methods. Chapters 4, 5, and 6 present the studies on the *k2 Mdh1-n y20* unstable chromosomal region. Chapter 4 is a manuscript ready to be submitted to *Journal of Heredity*. This paper presents genetic studies on four mutants newly found or generated in the *k2 Mdh1-n y20* chromosomal region. In this research, my participation included making crosses of Williams X T239 and T261, and verifying that the mutation with tan-saddle seeds found in cv. Kenwood was not caused by contamination. Chapter 5 is a manuscript submitted to *Theoretical and Applied Genetics*. This paper describes the work on mapping the *k2*, *Mdh1-n*, and *y20* loci with SSR markers using five mapping populations. Chapter 6 describes research on identifying the sizes and breakpoints of the deletions in six *Mdh1-n* or *Mdh1-n y20* mutants. Chapter 7 is a general conclusion that summarizes all the results of the studies.

All the paper or manuscripts presented in Chapters 2 to 6 had two authors, Min Xu and Dr. Reid G. Palmer. Dr. Reid G. Palmer, as my major professor, provided the guidance and directions throughout the entire research period. He also did major work described in Chapter 4 including the identification of mutations, allelism tests, and most of the cross-pollinations.

### **3. Literature review**

#### **3.1. Transposable elements in plant genomes**

Transposable elements are a class of DNA elements that can move from one genomic location to another. They were first discovered in maize by Barbara McClintock in the late 1940s, and later found ubiquitous in many organisms from bacteria to higher plants and animals (for review, see Feschotte et al. 2002). According to the intermediates they used for transposition, transposable elements could be classified into two major groups, RNA-mediated transposons and DNA-mediated transposons.

##### **3.1.1. RNA-mediated transposable elements**

The transposable elements using RNA as intermediates for transposition are generally termed retrotransposable elements or retrotransposons. During transposition, a retrotransposon first would be transcribed to RNA that could be reverse-transcribed into DNA that then integrates into the genome and generate new copies (for review, see Feschotte et al. 2002). Based on the structure organization, retrotransposons could be divided into two subgroups LTR (long terminal repeat) elements and non-LTR elements.

LTR elements have shown a strong structural similarity to retrovirus, for example, they have direct LTRs flanking the coding region, and encode *gag* and *pol* genes; however, usually, they do not contain the *envelope* (*env*) gene that encodes a transmembrane protein that is essential to retrovirus infection, so that they are not infectious (for review, see Grandbastien, 1992). The *pol* gene encodes a polyprotein responsible for four enzyme activities including protease, integrase, reverse transcriptase, and RNaseH. According to the constitution of the *pol* gene, LTR retrotransposons could be subdivided into two subclasses

*Ty1/copia* like elements and *Ty3/gypsy* like elements. In retrotransposons of the *Ty1/copia* group, regions in *pol* gene responsible for four different enzyme activities are arranged in this order: protease, integrase, reverse transcriptase, and RNaseH, while in retrotransposons of the *Ty3/gypsy* group, the region for integrase is located downstream of the encoding regions for protease, reverse transcriptase, and RNaseH (for review, see Casacuberta and Santiago 2003). Many elements were identified from these two groups in plants (Pouteau et al. 1991; Hirochika 1993; Hirochika et al. 1996; Vicient et al. 1999; Myers et al. 2001).

However, some elements of LTR retrotransposon were found encoding an *env*-like protein referred as endogenous retroviruses (Wright and Voytas 2002), for example, *Ty1/copia* element *Endovir* from *Arabidopsis* (Peterson-Burch et al. 2000), *Ty3/gypsy* element *Athila* from *Arabidopsis* (Wright and Voytas 1998), and *Ty3/gypsy* element *Cyclops-2* from pea (Chavanne et al. 1998).

Non-LTR retrotransposons lack of LTRs, and their transcription starts from internal promoters. Non-LTR retrotransposons could be subdivided to several groups. The first group is LINEs (long interspersed nuclear elements). LINEs are distinguished from others by an A-rich tail. It encodes three proteins including gag-like protein, endonuclease, and reverse transcriptase (for review, see Feschotte et al. 2002). The first isolated LINE in plants was *Cin4* element from maize (Schwarz-Sommer et al. 1987). The amino acid sequences of LINEs were conserved among 53 LINE homologues from 27 plant species including both monocots and dicots (Noma et al. 1999).

The second group in this category is SINEs (short interspersed nuclear elements) such as *TS* in tobacco (Yoshioka et al. 1993) and *S1* in *Brassica* (Deragon et al. 1994). Unlike LINEs, SINEs do not encode any enzyme activities. Some of them have a 3' tail homologous

to the one in LINEs, so they can be transposed by using enzymes from those LINEs that shared homologous 3' tail with them (Kajikawa and Okada 2002).

The copy number of retrotransposons varies among plant species. For example, the retrotransposons in *Arabidopsis*, which has relatively small genome size (~130 Mb), have very low copy number. *Tat1* (Wright and Voytas 1998) and *Athila* (Pelissier et al. 1995) are the only elements found in *Arabidopsis* with moderate numbers. And totally retrotransposons contributed 14% of the genome and tended to clustering in the gene-poor pericentromeric regions (The Arabidopsis Genome Initiative 2000).

However, in species with larger genomes such as maize, repetitive sequences (mostly retrotransposons) account for 50-80% of the genome (SanMiguel and Bennetzen 1998). Both LTR and non-LTR elements could reach very high copy number. In LTRs, *Ty1/copia* like element *Opie-2* in maize genome could have as many as 100,000 copies (Myers et al. 2001); BARE-1 from barley genome was present in 5000-22,000 copies (Vicent et al. 1999); and *Ty3/ gypsy* element *Huck-2* from maize had 200,000 copies (Meyers et al. 2001). In non-LTRs, LINE element *del2* from *Lilium* could reach 25,000 copies (Leeton and Smyth, 1993); SINE element TS from tobacco could reach 50,000 copies (Yoshioka et al. 1993).

In the plant kingdom, LTR elements are the most popular retrotransposons. They are interspersed in the whole genome. Sometimes, they can insert into a preexisted element and generate a nested retrotransposons (SanMiguel et al. 1996). Genomics study indicated that LTR retrotransposons in plants experienced a high level of LTR-retrotransposon amplification in recent times (less than 15 million years) (SanMiguel et al. 1998; Kumar and Bennetzen 1999; Devos et al. 2002); however, simultaneously, they are lost at a very fast speed. It was found that most of the detectable LTR retrotransposons in rice inserted less than



8 million years ago, and have now lost over two-thirds of their encoded sequences through deletion (Ma et al. 2004). In *Arabidopsis*, LTR-retrotransposons were mainly removed through illegitimate recombination (Devos et al. 2002), whereas in rice, LTR-retrotransposons were removed through both unequal homologous recombination and illegitimate recombination (Ma et al. 2004).

Although there are large numbers of retrotransposable elements present in plant genomes, but only a few families have been reported to have transposition activity under certain circumstances such as tissue culture, for example, *Tnt1* and *Tto1* of tobacco (Pouteau et al. 1991; Hirochika 1993), and *Tos17* of rice (Hirochika 2001). These elements are all present in genome at very low copies. By blasting to the *Triticeae* EST data base, 0.064% EST and 0.081% EST were found to be significantly similar to *Ty1/copia* and *Ty3/gypsy*, respectively (Echenique et al. 2002). The percentage found in cDNA libraries from plants under stress conditions was three- to four-fold higher than in cDNA libraries from plants grown under normal conditions. Recently, a *Ty1/copia* like element with about 400 copies was found active in callus of sweet potato by the 3' RACE technique (Tahara et al. 2004), which creates a new way to identify the active retrotransposons.

Transcriptional gene silencing (TGS) induced by high copy number was proposed to be responsible for low transcription activity of retrotransposons (for review, see Casacuberta and Santiago 2003). Most plant retrotransposons are highly methylated so that they could hardly be transcribed (Bennetzen et al. 1994; Hirochika 1997). The *ddm1* mutation that would reduce the DNA methylation could reactivate silenced the retrotransposon (Hirochika et al. 2000), which supported the hypothesis.

### 3.1.2. DNA-mediated transposons

DNA-mediated transposable elements make transposition using “cut and paste” mechanism. An obvious structural character of this class of transposons is they contain short or long terminal inverted repeats (TIR) (for review, see Feschotte et al. 2002). Only one or two genes encoding transposase are essential for transposition. During transposition, transposase recognizes the TIRs and cuts the transposon out; then the excised transposon would randomly insert back into the genome (for review, see Gierl and Saedler 1992). Transposons that encode functional transposases could transpose by themselves, and they are named autonomous transposable elements. In many cases, the transposase genes in transposons are truncated or deleted, which leads to transposons cannot transpose by themselves. They need transposase from its autonomous homologues to accomplished transposition. This kind of transposable elements is named non-autonomous transposable elements. Usually DNA-mediated transposons are present in plant genomes with very low copy number, while MITEs, recently grouped in to DNA-mediated transposons, have many copies in the genome (for review, see Casacuberta and Santiago 2003). According to the transposase and TIR similarity, plant DNA transposons can be divided to 5 superfamilies: *hAT*, *CACTA*, *Mutator*, *PIF/Harbinger*, and *Tc1/mariner* (for review, see Feschotte et al. 2002).

The first recognized transposable element system, *Ac/Ds*, is a representative of the *hAT* family. TIRs of *hAT* elements are usually short. Elements in this family share a conserved domain of ~50 amino acids located at the C terminus, and they would generate a 8-bp host duplication during insertion (for review, see Rubin et al. 2001).

*Spm (En)/dspm* was the second transposable element system found in plants. They are different with *AC/DS* in structure. They have TIRs terminated with a conserved motif CACTA, so that the elements with similar structure are also referred to CACTA elements. After excision, CACTA elements would form a 3-bp host duplication upon insertion (for review, see Grandbastien, 1992). Recently, it was found CACTA transposons could reach very high copies in *Triticeae*, but these high-copied elements could not produce functional transposase, and belong to non-autonomous elements (Wicker et al. 2003).

*Mutator (MuDR/Mu element)* was also first isolated from maize, and later *Mu*-like elements were found in rice and *Arabidopsis* (for review, see Lisch 2002). All maize *Mu* elements contain conserved relative long TIRs. The autonomous element *MuDR* has two genes in coding region, *mudrA* that encode a 120 kDa transposase (MURA) and *mudrB* that encodes a 23 kDa protein (MURB).

*PIF/Harbinger* elements such as *Tourist* (Bureau and Wessier 1992) and *Tc1/mariner* elements such as *Stowaway* (Bureau and Wessier 1994) both belong to miniature inverted-repeat transposable elements (MITEs). They are very special transposable elements with characteristics similar to both RNA-mediated and DNA-mediated transposons (for review, see Feschotte et al. 2002; Casacuberta and Santiago 2003). For example, they have TIRs, which is similar to the defective DNA-mediated transposons, but their copy number is very high, which is similar to retrotransposable elements. For example, *Stowaway* elements in *Tc1/mariner* family constitute 2% of the rice genome (Mao et al. 2000). However, other characteristics of MITEs such as small size (usually less than 500 bp), AT-rich sequence, and target site preference distinguish them from other transposable elements (for review, see Feschotte et al. 2002; Casacuberta and Santiago 2003).

As to transposition activity, normally, DNA-mediated transposons are very active. The transposition frequencies are greater than  $10^{-2}$  germline excisions per generation (Flavell et al. 1994). In the most active transposon family, *Mu* elements, transposition frequency can reach 100% (an average of one transposition event per element per generation) (for review, see Lisch 2002). Unlike most of DNA-mediated transposons, the transposition activities of MITEs are almost undetectable. So far, only one autonomous element was identified, which is the rice *mPing* element (Jiang et al. 2003).

### **3.2. Transposable elements as mutagens**

Transposable elements are well known for their ability to generate both small-scale mutations and large-scale chromosomal rearrangements such as inversions, duplications, and deletions. *Ac/Ds* elements of maize first were recognized by studying the phenomenon that *Ds* elements could induce chromosome breakage, form “breakage-fusion-bridge-cycle”, and lead to gross chromosome rearrangements (for review, see Lönnig and Saedler 2002).

#### **3.2.1. Small-scale mutations**

One kind of small-scale mutation is caused by transposon insertion. Both retrotransposons and DNA transposons can lead to mutations of genes where they insert. In *niv<sup>rec</sup>:98* allele of snapdragon, insertion of *Tam3* in the promoter region of *nivea* gene could lead to pale pigmented flowers; somatic excision of *Tam3* would recover the gene expression and lead to dark red sectors or stripes; hence, the mutant with *niv<sup>rec</sup>:98* allele produces variegated flowers (Sommer et al. 1985). Not only insertion in the regulatory region or exons of a gene would affect this gene’s expression, insertion in introns sometimes also would generate mutations. For example, the *CsI* element of sorghum causes the *Y* gene mutation by

inserting into its second intron (Chopra et al. 1999). Many transposons were isolated by cloning the gene they mutate (for review, see Casacuberta and Santiago 2003). For example, the maize retrotransposon *Bs1* was found in an insertional mutation of the *Adh* gene (Johns et al. 1985).

The second kind of mutation is caused by transposon imprecise excision, which would leave long or short footprints that could cause frameshifts, or generate small deletions adjacent to the excision site. For example, imprecise excision of *Tam3* element from *pallida* locus generated a series of deletions nearby and led to various spatial patterns of flower color (Almeida et al. 1989).

### 3.2.2. Large-scale mutations

Large deletions could be induced by two linked transposons. In maize, *P1-rr11* allele at *p1* locus contains two transposable elements *Ac* and *fAc* (*fractured Ac*) with ~13kb apart and same orientation. Mutants with ~13kb deletion were found in its progenies, which resulted from transposition of a piece of DNA involving the 5' end of an intact *Ac* element and the 3' end of *fAc* (Zhang and Peterson 2004). The *nev*:592 mutant in snapdragon is due to a large size chromosomal deletion caused by *Tam3* insertion, while the deletion doesn't include the *Tam3* element (Martin and Lister 1989). The deletion region is from the 3'-end TIR of *Tam3* to a 11bp sequence that shows strong homology to the TIR of *Tam3*. One hypothesis is that transposase could recognize this region as a transposon, cut it out, and cause a large deletion. The similar situations were reported in *Arabidopsis* when using *Ac/Ds* enhancer detection and gene-trap system generated mutations (Page et al. 2004). Large deletions adjacent to *Ds* with size ranging from 64 to 104 kb were identified. It was hypothesized that deletions were generated when transposase recognized the end of a newly

inserted *Ds* element and the end of pre-existed *Ds* element at the donor site as two ends of one transposon and excised out the region between these two ends.

In addition to deletions, transposable elements would also induce inversions and duplications. For example, *niv<sup>rec</sup>*:531 derived from *niv<sup>rec</sup>*:98 is an inversion mutant with size greater than 20 kb (Martin and Lister 1989). Another two alleles *niv<sup>rec</sup>*:554 and *niv<sup>rec</sup>*:557 are inverted duplications with size of 4.3 kb and 2.3 kb, respectively (Martin and Lister 1989).

### 3.3. Transposable elements in functional genomics

With many genome sequencing projects progressing, more and more genes are sequenced in various plant species. Currently, the whole genomes of *Arabidopsis* (~130 Mb) and rice (~430 Mb) are sequenced, but only a few of them are known functionally. So, the aim of functional genomics is to identify functions of all the genes. To reach the goal, generating mutants for each gene is the first step.

So far, major methods of mutagenesis are T-DNA and TE insertion (for review, see Parinov and Sundaresan 2000; Walbot 2000), RNAi and anti-sense knockout (Wesley et al. 2001), TILLING (Targeting Induced Local Lesions IN Genomes) in which mutations are generated with EMS (McCallum et al. 2000), and ionizing radiation such as fast neutron deletion (Li et al. 2001).

TEs mutagenesis has many advantages. First, it can be used in both forward and reverse genetics research, whereas methods of RNAi and antisense knockout, TILLING, and fast neutron deletion usually are designed for reverse genetics research. In forward genetics, transposon tagging technology is well developed and used widely to isolate genes from maize and snapdragon (for review, see Gierl and Saedler 1992). The first step of this method is to identify the mutation caused by transposon insertion. Subsequently, the gene can be isolated

by cloning the DNA fragments flanking the transposons. This method is helpful for cloning genes without knowing sequences and the nature of their product. So it is very useful in genomics study of those species whose genome sequences are not well defined. In reverse genetics, transposable elements could generate many mutant lines. The lines can be screened by PCR using a specific primer from a gene of interest and a primer from transposable element. Specific PCR products can then be tested for association with the mutant phenotype(s).

Second, the phenotype caused by transposon insertion is easier to confirm than the one caused by other mutagenesis methods. Transposons would excise out from the disrupted gene they inserted in the presence of transposase so that the mutation they caused could be reversed somatically and germinally (for review, see Parinov et al. 2000; Walbot 2000).

Third, transposon insertion sites are “hot spot” for secondary mutation. Particularly, *Ac/Ds* elements preferentially undergo local transposition, and those transposons integrated nearby would remobilize back to the original inserted site and cause many new alleles (for review, see Parinov et al. 2000; Walbot 2000).

*Ac/Ds* gene trap system using T-DNA as mediate was successfully generated for functional genomics studies in rice (Kolesnik et al. 2004) and *Arabidopsis* (for review, see Springer 2000). In this method, the *Ds* element harbors a *GUS* reporter gene with minimal promoter (enhancer trap) or without promoter (gene trap). Thus, the reporter gene would only be expressed when it inserted into a region near an enhancer (enhancer trap) or downstream of a promoter (promoter trap). The inserted *Ds* element in the mutant line could be subsequently mobilized by crossing to a line with *Ac* elements, and generate somatic and germinal reversions (for review, see Springer 2000).

By using *Ac/Ds* elements and *Cre-lox* site-specific recombination system, a mutagenesis system for generating large chromosome rearrangements was developed (Medberry et al. 1995; Osborne et al. 1995). The T-DNA harbors two *lox* loci, one of which resides in the *Ds* element and forms a *Ds-lox* complex. When the insertion lines produced by T-DNA transformation are crossed to the *Ac* homologous lines, transposase could move the *Ds-lox* complex to a nearby region. Then, deletion or inversion between two *lox* loci could be generated by introducing *Cre* recombinase through crossing.

This method was simplified and called TREGED (transposon-and recombinase-mediated genome deletion) (Liu et al. 2004). The TREGED construction consists of the maize *Ac/Ds* transposon, the yeast *R-RS* site-specific recombination system, the bacterial *tetR* repression systems, a novel artificial superintron, and the marker genes *GUS* and *Lc*. Deletions could be generated in the insertion line through only one cross that introduces transposase into this line.

### 3.4. Transposable element studies in soybean

The soybean genome contains 40-60% repetitive sequences (Goldberg 1978; Gurley et al. 1979), but how many are transposable elements remains unknown. So far, both retrotransposons and DNA-mediated transposons have been discovered in the soybean genome. However, only LTR retrotransposons and non-autonomous DNA-mediated transposons were discovered.

Both types of LTR elements, *Ty1/copia* and *Ty3/gypsy*, were found in the soybean genome (Bhattacharyya et al. 1997; Laten et al. 1998; Wright and Voytas 2002; Yano et al. 2005). Except for the *copia*-like *Tgmr* retrotransposon identified by Bhattacharyya et al. (1997), all the elements are retrovirus-like elements (endogenous retroviruses). The *SIRE1*



family is *Ty1/copia* like element (Laten et al. 1998). By examining eight elements in this family, it was suggested that they are relatively young. Two of them were transposed within 70,000 years. Southern blot showed that there are more than 1000 copies of them in the soybean genome. DNA sequences of these eight examined elements showed 95% nucleotide identity (Laten et al. 2003). *SIRE1* elements are 11kb in length (Laten et al. 1998). The *gag/pol* and *env*-like genes are in the same reading frame but separated by a single UAG stop codon. The *env*-like gene is expressed using a stop codon suppression mechanism (Havecker and Voytas 2003).

*Calypso* (Wright and Voytas 2002) and *Diaspora* (Yano et al. 2005) families belong to *Ty3/gypsy* elements; both of them are highly degenerated. *Calypso* elements have sizes ranging from 12-14 kb and LTRs with sizes ranging from 1.3-1.8 kb (Wright and Voytas 2002). They contain an 1801-codon ORF (open reading frame) for a *gag-pol* polyprotein, and an *env*-like ORF. *Diaspora* elements are 11,737 bp in length and contain a single 1892-codon ORF encoding a *gag-pol* polyprotein. Although *Diaspora* elements don't contain the *env* gene, phlogenetic analysis implied that they are closely related to the *Athila* and *Calypso* retroelements from *Arabidopsis* and soybean, respectively. Thus they could have evolved from an endogenous retrovirus.

Fragments of *Calypso*-like elements, a recently inserted *SIRE1* element, and a *SIRE1* solo-LTR were found in a BAC clone corresponding to the pericentromeric regions of soybean linkage group L. FISH (fluorescence *in situ* hybridization) analysis showed the distribution of these retroelements was associated with heterochromatic regions of the soybean chromosome (Lin et al. 2005).

Two types of DNA-mediated transposons were identified in soybean. Seven members of *Tgm* families were isolated with sizes range from 1.6-12 kb, which were caused by different internal deletions (Rhodes and Vodkin 1988). The elements were flanked with 54 bp TIRs. The deduced amino acid sequences from ORFs isolated from *Tgm* 4 and 5 are similar to the ones from a portion of ORF1 from *En-1* element of maize, which places them into the CACTA element group.

The other one is *Soymar1* family, which belongs to *Tc1/mariner*-like elements (Jarvik and Lark 1998). While the *Soymar1* elements had different TIRs to other *mariner*-like elements, they do have the "D35E" motif specific for *mariner* elements, that is, the glutamic acid residue of the "D35E" is replaced by a second aspartic acid residue.

Since soybean transformation efficiency is relatively low, insertional mutagenesis using heterogeneous transposable elements or T-DNA is not suitable for soybean functional genomics research. Therefore, searching and identifying autonomous transposable elements are very important.

So far, several mutable alleles have been reported in soybean, for example, *w4-m* and *wp-m* mutable alleles that condition variegated flowers (Palmer et al. 1990; Johnson et al. 1998), *r-m* mutable allele that makes seed with variegated seed coat (Chandlee and Vodkin 1989a), and *Y18-m* mutable allele that conditions variegated foliage (Chandlee and Vodkin 1989b). All of them could generate somatic and germinal reversions which are similar to the phenotype caused by transposon movement. An unstable chromosomal region consisting of *k2*, *Mdh1-n*, and *y20* genes was also suspected to harbor a non-autonomous transposon (Chen and Palmer 1998b). Further studies to the *w4-m* allele and the *k2 Mdh1-n y20* unstable chromosomal region are presented in this dissertation.

#### 4. References

- Almeida J, Carpenter R, Robbins TP, Martin C, Coen ES (1989) Genetic interactions underlying flower color patterns in *Antirrhinum majus*. *Genes Dev* 3:1758-1767
- Amberger LA, Shoemaker RC, Palmer RG (1992) Inheritance of two independent isozyme variants in soybean plants derived from tissue culture. *Theor Appl Genet* 84: 600-607
- Bhattacharyya MK, Gonzales RA, Kraft M, Buzzell RI (1997) A *copia*-like retrotransposon *Tgmr* closely linked to the *Rps1-k* allele that confers race-specific resistance of soybean to *Phytophthora sojae*. *Plant Mol Biol* 34: 255-264
- Bennetzen JL, Schrick K, Springer PS, Brown WE, SanMiguel P (1994) Active maize genes are unmodified and flanked by diverse classes of modified, highly repetitive DNA. *Genome* 37: 565-576
- Bureau TE, Wessler SR (1992) *Tourist*: a large family of small inverted repeat elements frequently associated with maize genes. *Plant Cell* 41: 283-294
- Bureau TE, Wessler SR (1994) *Stowaway*: a new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. *Plant Cell* 6: 907-916
- Casacuberta JM, Santiago N (2003) Plant LTR-retrotransposons and MITEs: control of transposition and impact on the evolution of plant genes and genomes. *Gene* 311: 1-11
- Chandlee JM, Vodkin LO (1989a) Unstable expression of a soybean gene during seed coat development. *Theor Appl Genet* 77: 587-594
- Chandlee JM, Vodkin LO (1989b) Unstable genes affecting chloroplast development in soybean. *Dev Genet* 10: 532-541

- Chavanne F, Zhang DX, Liaud MF, Cerff R (1998) Structure and evolution of *Cyclops*: a novel giant retrotransposon of the *Ty3/Gypsy* family highly amplified in pea and other legume species. *Plant Mol Biol* 37: 363-375
- Chen XF, Imsande J, Palmer RG (1999) Eight new mutants at the *k2 Mdh1-n y20* chromosomal region in soybean. *J Hered* 90: 399-403
- Chen XF, Palmer RG (1996) Inheritance and linkage with the *k2* and *Mdh1-n* loci in soybean. *J Hered* 87: 433-347
- Chen XF, Palmer RG (1998a) Recombination and linkage estimation between the *k2* and *Mdh1-n y20* loci in soybean. *J Hered* 89: 488-494
- Chen XF, Palmer RG (1998b) Instability at the *k2 Mdh1-n y20* chromosomal region in soybean. *Mol Gen Genet* 260: 309-318
- Chopra S, Brendel V, Zhang J, Axtell JD, Peterson T (1999) Molecular characterization of a mutable pigmentation phenotype and isolation of the first active transposable element from *Sorghum bicolor*. *Proc Natl Acad Sci USA* 96: 15330-15335
- Deragon JM, Landry BS, Pélissier T, Tutois S, Tourmente S, Picard G (1994) An analysis of retroposition in plants based on a family of SINEs from *Brassica napus*. *J Mol Evol* 39: 378-386
- Devos KM, Brown JKM, Bennetzen JL (2002) Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Res* 12: 1075-1079
- Echenique V, Stamova B, Wolters P, Lazo G, Carollo L, Dubcovsky J (2002) Frequencies of *Ty1-copia* and *Ty3-gypsy* retroelements within the *Triticeae* EST databases. *Theor Appl Genet* 104: 840-844

- Feschotte C, Jiang N, Wessler SR (2002) Plant transposable elements: where genetics meets genomics. *Nature Rev Genetics* 3: 329-341
- Flavell AJ, Pearce SR, Kumar A (1994) Plant transposable elements and the genome. *Curr Opin Genet Dev* 4: 838-844
- Gierl A, Saedler H (1992) Plant transposable elements and gene tagging. *Plant Mol Biol* 19: 39-49
- Goldberg RB (1978) DNA sequence organization in the soybean plant. *Biochem Genet* 16: 45-68
- Grandbastien MA (1992) Retroelements in higher plants. *Trends Genet* 8: 103-108
- Groose RW, Palmer RG (1991) Gene action governing anthocyanin pigmentation in soybean. *J Hered* 82: 498-501
- Groose RW, Schulte SM, Palmer RG (1990) Germinal reversion of an unstable mutation for anthocyanin pigmentation in soybean. *Theor Appl Genet* 79: 161-167
- Groose RW, Weigelt HD, Palmer RG (1988) Somatic analysis of an unstable mutation for anthocyanin pigmentation in soybean. *J Hered* 79: 263-267
- Gurley WB, Hepburn AG, Key JL (1979) Sequence organization of the soybean genome. *Biochim Biophys Acta* 561: 167-183
- Havecker ER, Voytas DF (2003) The soybean retroelement *SIRE1* uses stop codon suppression to express its envelope-like protein. *EMBO Rep* 4: 274-277
- Hirochika H (1993) Activation of tobacco retrotransposons during tissue culture. *EMBO J* 12: 2521-2528
- Hirochika H (2001) Contribution of the *Tos17* retrotransposon to rice functional genomics. *Curr Opin Plant Biol* 4: 118-122

- Hirochika H, Okamoto H, Kakutani T (2000) Silencing of retrotransposons in *Arabidopsis* and reactivation by the *ddm1* mutation. *Plant Cell* 12: 357-369
- Hirochika H, Sugimoto K, Otsuki Y, Tsugawa H, Kanda M (1996) Retrotransposons of rice involved in mutations induced by tissue culture. *Proc Natl Acad Sci USA* 93: 7783-7788
- Hirochika H (1997) Retrotransposons of rice: Their regulation and use for genome analysis. *Plant Mol Biol* 35: 231-240
- Imsande J, Pittig J, Palmer RG, Wimmer C, Gietl C (2001) Independent spontaneous mitochondrial malate dehydrogenase null mutants in soybean are the result of deletions. *J Hered* 92: 333-338
- Jarvik T, Lark KG (1998) Characterization of *Soymar1*, a *mariner* element in soybean. *Genetics* 149: 1569-1574
- Jiang N, Bao Z, Zhang X, Hirochika H, Eddy SR, McCouch SR, Wessler SR (2003) An active DNA transposon family in rice. *Nature* 421: 163-167
- Johns MA, Mottinger J, Freeling M (1985) A low copy number, *copia*-like transposon in maize. *EMBO Journal* 4: 1093-1102
- Johnson EOC, Stephens PA, Fasoula DA, Nickell CD, Vodkin LO (1998) Instability of a novel multicolored flower trait in inbred and outcrossed soybean lines. *J Hered* 89: 508-515
- Kajikawa M, Okada N (2002) LINEs mobilize SINEs in the eel through a shared 3' sequence. *Cell* 111: 433-444
- Kato KK, Palmer RG (2003) Molecular mapping of male-sterile, female-sterile mutant gene (*st8*) in soybean. *J Hered* 94: 425-428

- Kato KK, Palmer RG (2004) Molecular mapping of four ovule lethal mutants in soybean. *Theor Appl Genet* 108: 577-585
- Kolesnik T, Szeverenyi I, Bachmann D, Kumar CS, Jiang S, Ramamoorthy R, Cai M, Ma ZG, Sundaresan V, Ramachandran S (2004) Establishing an efficient *Ac/Ds* tagging system in rice: large-scale analysis of *Ds* flanking sequences. *Plant J* 37: 301-314
- Kumar A, Bennetzen JB (1999) Plant retrotransposons. *Annu Rev Genet* 33: 479-532
- Laten HM, Majumdar A, Gaucher EA (1998) *SIRE-1*, a *copia*/Ty1-like retroelement from soybean, encodes a retroviral envelope-like protein. *Proc Natl Acad Sci USA* 95: 6897-6902
- Laten HM, Havecker ER, Farmer LM, Voytas DF (2003) *SIRE-1*, an endogenous family from *Glycine max*, is highly homogenous and evolutionary young. *Mol Biol Evol* 20: 1-13
- Leeton PR, Smyth DR (1993) An abundant LINE-like element amplified in the genome of *Lilium speciosum*. *Mol Gen Genet* 237: 97-104
- Li X, Song Y, Century K, Straight S, Ronald P, Dong X, Lassner M, Zhang Y (2001) A fast neutron deletion mutagenesis-based reverse genetics system for plants. *Plant J* 27: 235-242
- Lin JY, Hass-Jacobus B, Sanmiguel P, Walling JG, Yuan Y, Shoemaker RC, Young ND, Jackson S (2005) Centric regions of soybean (*Glycine max* L. Merr.) chromosomes consist of retroelements and tandemly repeated DNA and are structurally and evolutionarily labile. *Genetics* (In press)
- Lisch D (2002) Mutator transposons. *Trends Plant Sci* 7: 498-504

- Liu Z, Davies CS, and Mount DW (2000) TREGED: A new strategy for inducing deletions in plant genomes. *Plant Mol Biol Rep* 18: 255-263
- Lönnig WE, Saedler H (2002) Chromosome rearrangements and transposable elements. *Annu Rev Genet* 36: 389-410
- Ma J, Devos KM, Bennetzen JL (2004) Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. *Genome Res* 14: 860-869
- Mao L, Wood TC, Yu Y, Budiman MA, Tomkins J, Woo S, Sasinowski M, Presting G, Frisch D, Goff S, Dean RA, Wing RA (2000) Rice transposable elements: a survey of 73,000 sequence-tagged-connectors. *Genome Res* 10: 982-990
- Martin C, Lister C (1989) Genome juggling by transposons: Tam3-induced rearrangements in *Antirrhinum majus*. *Dev Genet* 10: 438-451
- McCallum CM, Comai L, Greene EA, Henikoff S (2000) Targeting induced local lesions IN genomes (TILLING) for plant functional genomics. *Plant Physiology* 123: 439-442
- Medberry SL, Dale E, Qin M, Ow DW (1995) Intra-chromosomal rearrangements generated by *Cre-lox* site-specific recombination. *Nucleic Acids Res* 23: 485-490
- Meyers BC, Tingey SV, Morgante M (2001) Abundance, distribution and transcriptional activity of repetitive elements in the maize genome. *Genome Res* 11: 1660-1676
- Noma K, Ohtsubo E, Ohtsubo H (1999) Non-LTR retrotransposons (LINEs) as ubiquitous components of plant genomes. *Mol Gen Genet* 261: 71-79
- Osborne BI, Wirtz U, Baker B (1995) A system for insertional mutagenesis and chromosomal rearrangement using the *Ds* transposon and *Cre-lox*. *Plant J* 7: 687-701



- Page DR, Kohler C, Da Costa-Nunes JA, Baroux C, and Moore JM (2004) Intrachromosomal excision of a hybrid *Ds* element induces large genomic deletions in *Arabidopsis*. *Proc Natl Acad Sci USA* 101: 2969-2974
- Palmer RG, Burzlaff JD, Shoemaker RC (2000) Genetic analyses of two independent chlorophyll-deficient mutants identified among the progeny of a single chimeric foliage soybean plant. *J Hered* 91: 297-303
- Palmer RG, Pfeiffer TW, Buss GR, Kilen TC (2004) Qualitative genetics. In Boerma HR, Specht JE (eds) *Soybeans: Improvement, Production, and Uses*, 3rd edn. ASA, CSSA, SSSA, Madison, WI, pp 137-214
- Palmer RG, Groose RW (1993) A new allele at the *w4* locus derived from the *w4-m* mutable allele in soybean. *J Hered* 84: 297-300
- Palmer RG, Groose RW, Weigelt HD, Miller JE (1990) Registration of a genetic stock (*w4-mw4-m*) for unstable anthocyanin pigmentation in soybean. *Crop Sci* 30: 1376-1377
- Palmer RG, Hedges BR, Benavente RS, Groose RW (1989) *w4*-mutable line in soybean. *Dev Genet* 10: 542-551
- Palmer RG (1984) Pleiotropy or close linkage of two mutants in soybeans. *J Hered* 75: 457-462
- Parinov S, Sundaresan V (2000) Functional genomics in *Arabidopsis*: large-scale insertional mutagenesis complements the genome sequencing project. *Curr Opin Biotechnol* 11: 157-161
- Pelissier T, Tutois S, Deragon JM, Tourmente S, Genestier S, Picard G (1995) *Athila*, a new retroelement from *Arabidopsis thaliana*. *Plant Mol Biol* 29: 441-452

Peterson-Burch BD, Wright DA, Laten HM, Voytas DF (2000) Retroviruses in plants?

Trends Genet 16: 151-152

Pouteau S, Huttner E, Grandbastien MA, Caboche M (1991) Specific expression of the

tobacco *Tnt1* retrotransposon in protoplasts. EMBO J 10: 1911-1918

Rhodes PR, Vodkin LO (1988) Organization of the Tgm family of transposable elements in

soybean. Genetics 120: 597-604

Rubin E, Lithwick G, Levy AA (2001) Structure and evolution of the *hAT* transposon

superfamily. Genetics 158: 949-957

SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, Melake-Berhan A,

Springer PS, Edwards KJ, Lee M, Avramova Z, Bennetzen JL (1996) Nested

retrotransposons in the intergenic regions of the maize genome. Science 274: 765-768

SanMiguel P, Bennetzen JL (1998) Evidence that a recent increase in maize genome size was

caused by the massive amplification of intergene retrotransposons. Ann Bot 82: 37-44

SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL (1998) The paleontology of

intergene retrotransposons of maize. Nat Genet 20: 43-45

Schwarz-Sommer Z, Leclercq L, Goebel E, Saedler H (1987) *Cin4*, an insert altering the

structure of the *Al* gene in *Zea mays*, exhibits properties of nonviral retrotransposons.

EMBO J 6: 3873-3880

Sommer H, Carpenter R, Harrison BJ, Saedler H (1985) The transposable element Tam3 of

*Antirrhinum majus* generates a novel type of sequence alterations upon excision. Mol

Gen Genet 199: 225-231

Springer PS (2000) Gene traps: tools for plant development and genomics. Plant Cell 12:

1007-1020

- Tahara M, Aoki T, Suzuka S, Yamashita H, Tanaka M, Matsunaga S, Kokumai S (2004)  
Isolation of an active element from a high-copy-number family of retrotransposons in the sweetpotato genome. *Mol Genet Gen* 272: 116-27
- The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796-815
- Vicient CM, Suoniemi A, Ananthawat-Jonsson K, Tanskanen J, Beharav A, Nevo E, Schulman AH (1999) Retrotransposon *BARE-1* and its role in genome evolution in the genus *Hordeum*. *Plant Cell* 11: 1769-1784
- Walbot V (2000) Saturation mutagenesis using maize transposons. *Curr Opin Plant Biol* 3: 103-107
- Wesley SV, Helliwell CA, Smith NA, Wang MB, Rouse DT, Liu Q, Gooding PS, Singh SP, Abbott D, Stoutjesdijk PA (2001) Construct design for efficient, effective and high-throughput gene silencing in plants. *Plant J* 27: 581-590
- Wicker T, Guyot R, Yahiaoui N, Keller B (2003) CACTA transposons in *Triticeae*. A diverse family of high-copy repetitive elements. *Plant Physiol* 132: 52-63
- Wright DA, Voytas DF (2002) *Athila4* of *Arabidopsis* and *Calypso* of soybean define a lineage of endogenous plant retroviruses. *Genome Res* 12: 122-131
- Wright DA, Voytas DF (1998) Potential retroviruses in plants: *Tat1* is related to a group of *Arabidopsis thaliana* *Ty3/gypsy* retrotransposons that encode envelope-like proteins. *Genetics* 149: 703-715
- Yano ST, Panbehi B, Das A, Laten HM (2005) *Diaspora*, a large family of *Ty3-gypsy* retrotransposons in *Glycine max*, is an envelope-less member of an endogenous plant retrovirus lineage. *BMC Evol Biol* (in press)

- Yoshioka Y, Matsumoto S, Kojima S, Ohshima K, Okada N, Machida Y (1993) Molecular characterization of a short interspersed repetitive element from tobacco that exhibits sequence homology to specific tRNAs. *Proc Natl Acad Sci USA* 90: 6562-6566
- Zhang J, Peterson T (2004) Transposition of reversed *Ac* element ends generates chromosome rearrangements in maize. *Genetics* 167: 1929-1937

## CHAPTER 2 GENETIC ANALYSIS AND MOLECULAR MAPPING OF A PALE FLOWER ALLELE AT THE *W4* LOCUS IN SOYBEAN

A paper published in *Genome* 2005, 48: 334-340

Min Xu and Reid G. Palmer

### Abstract

In soybean [*Glycine max* (L.) Merr.], the *w4*-mutable line that harbors the *w4-m* allele was identified in 1983. It was proposed that this line contained an autonomous transposable element at the *W4* locus, which is a major locus controlling the biosynthesis of anthocyanin. The *w4-m* allele can revert to the *W4* allele that produces the wild-type phenotype, or sometimes to other alleles that produce intermediate phenotypes. Mutant plants that produce pale flowers were identified among the progeny of a single germinal revertant event from the *w4*-mutable line. Through genetic analysis, we established that the pale-flower mutation was conditioned by a new allele (*w4-p*) at the *W4* locus. The *w4-p* allele is dominant to the *w4* allele but recessive to the *W4* allele; and the *w1* allele has an epistatic effect on the *w4-p* allele. The pale-mutant line (*w4-pw4-p*) was designated as Genetic Type Collection number T369. An F<sub>2</sub> mapping population derived from the cross of Minsoy (*W4W4*) x T369 (*w4-pw4-p*) was used to map the *W4/w4-p* locus by using simple sequence repeat (SSR) markers. The *W4* locus was located at one end of molecular linkage group D2, 2.3 cM from the SSR marker Satt386 and close to the nearby telomere.

## Introduction

In soybean, 6 genes, *W1*, *W2*, *W3*, *W4*, *Wp*, and *Wm*, have been identified that condition anthocyanin pigmentation in flowers and hypocotyls (Hartwig and Hinson 1962; Stephens and Nickell 1992; Palmer et al. 2004). The *W4* locus has a major effect on the biosynthesis of anthocyanin in soybean flowers and hypocotyls. Mutations at the *W4* locus lead to less anthocyanin pigmentation in both tissues. In the *W1*\_ genetic background, soybean lines with genotype *W4*\_ generate wild-type purple flowers and purple hypocotyls. Soybean lines with genotype *w4w4* (a stable recessive mutation at the *W4* locus) generate near-white flowers and green hypocotyls (Groose and Palmer 1991).

A mutable allele (*w4-m*) that conditions unstable anthocyanin pigmentation in soybean flowers and hypocotyls was identified at the *W4* locus. The genetic stock for the inbred mutant line (*w4-mw4-m*) was registered as the *w4*-mutable line, and was assigned Genetic Type Collection number T322 (Palmer et al. 1990). It has been proposed that the *w4*-mutable line was proposed to contain an autonomous transposable element at or near the *W4* locus, which results in the *w4-m* allele (Palmer et al. 1989). The *w4-m* allele can revert to the *W4* allele after excision of the putative transposable element. Somatic reversion of the *w4-m* allele results in plants that produce variegated flowers and hypocotyls (Groose et al. 1988). Germinal reversion of the *w4-m* allele, in most cases, results in plants that produce all wild-type purple flowers and hypocotyls (Groose et al. 1990). Many new mutations were identified among self-pollinated progenies of these wild-type germinal revertants, for example, chlorophyll-deficient leaves, necrotic roots, and complete or partially sterile plants (Palmer et al. 1989). Occasionally, germinal revertants may produce intermediate colored flowers (Groose et al. 1990).

Two mutant lines with intermediate colored flowers (dilute purple and pale) have been identified from the *w4*-mutable lines. The dilute-purple-mutant line was conditioned by a new allele at the *W4* locus (*w4-dp*), and was assigned Genetic Type Collection number T321 (Palmer and Groose 1993).

Genetic analysis of the pale-mutant line has not been reported. The pale-mutant line produces pale flowers, which are lighter than wild-type purple but darker than recessive near-white ones (Fig. 1), and green hypocotyls. Genetic analysis and molecular mapping of the pale mutant should provide additional evidence to support the hypothesis that there is an active transposable element within the *w4-m* allele. If it is an allele at the *W4* locus, the pale-mutant line will provide new material that can be used to study gene expression and regulation of the *W4* gene.

In the soybean genome, simple sequence repeat (SSR) markers are highly polymorphic (Akkaya et al. 1992), randomly distributed on chromosomes (Akkaya et al. 1995), and suitable for mapping new genes because 1 pair of primers usually amplifies only 1 product from an inbred genotype with PCR (Cregan et al. 1994). An integrated genetic linkage map with 20 molecular linkage groups, including 606 SSR, 689 RFLP (restriction fragment length polymorphism), 79 RAPD (random amplified polymorphic DNA), 11 AFLP (amplified fragment length polymorphism), 10 isozyme, and 26 classical loci, has been constructed (Cregan et al. 1999). Of the six genes that control anthocyanin biosynthesis in soybean flowers, only two loci have been positioned on this map. The *W1* locus was located on molecular linkage group (MLG) F (Cregan et al. 1999) and the *wp* locus was located on MLG D1b+w (Hegstad et al. 2000).

Our objectives were to study the genetic relationship between the new pale mutation and the *W4* locus, and to position the *W4* locus on a soybean chromosome with SSR markers.

## **Materials and Methods**

### **Genetic experiments**

The pale-mutant line was determined to be true breeding. It originated as a revertant from the original *w4*-mutable line (Asgrow XP2214) (Palmer et al. 1997). Based upon the results in this paper, the Soybean Genetics Committee assigned gene symbol *w4-p* to the pale allele, and Genetic Type Collection number T369 to the pale-mutant line.

Minsoy (PI 27890), Harosoy *w4* isoline (L72-1138), and Harosoy *w1* isoline (L62-906) were each crossed with T369. The phenotypes and genotypes of these lines are described in Table 1. This was done to determine how many genes controlled the pale mutation, and to identify the relationship between the pale mutation and the *w4* locus and between the pale mutation and the *w1* locus. Progenies from all 3 crosses were analyzed in the F<sub>1</sub>, F<sub>2</sub>, and F<sub>3</sub> generations.

### **Molecular mapping experiment**

#### ***Plant materials***

The following true-breeding soybean lines were used in this experiment: Harosoy (purple flowers), Harosoy *w4* isoline (near-white flowers), Minsoy (purple flowers), and T369 (pale flowers) (Table 1).



The mapping population consisting of 114  $F_2$  individuals descended from a single  $F_1$  seed of Minsoy (purple flowers) x T369 (pale flowers) was grown at the Bruner Farm, near Ames, Iowa, in the summer of 2002. Plants were sampled, identified and threshed individually. Twenty-four  $F_{2:3}$  descendants from each  $F_2$  plant were planted for the progeny test in the fall of 2002 in the USDA greenhouse on the Iowa State University campus to determine the genotype of each  $F_2$  plant.

#### ***DNA extraction and PCR conditions***

DNA was extracted from freeze-dried young leaves of the Harosoy *w4* isolate, Minsoy, T369, and 114  $F_2$  plants of the mapping population by using the CTAB (cetyltrimethyl ammonium bromide) method (Keim et al. 1988).

PCRs were performed in 30  $\mu$ L of reaction mixture containing 1x PCR buffer, 1.75 mmol/L  $MgCl_2$ , 150  $\mu$ mol/L dNTP, 0.15  $\mu$ mol/L primers, 50 ng genomic DNA, and 3 units *Taq* DNA polymerase (Promega, Madison, Wis.). The PCR condition was 94 °C for 45 seconds, 47 °C for 45 seconds, and 68 °C for 45 seconds for a total of 32 cycles. The PCR products were evaluated by electrophoresis on 2% Agarose 3:1 (AMRESCO, Solon, Ohio) gels in 1x TBE (Tris/borate /EDTA) buffer, or on 8% polyacrylamide denaturing gels in 1x TAE (Tris/acetate/EDTA) buffer.

#### ***Denaturing polyacrylamide gel electrophoresis (PAGE) analysis***

PCR reactions were stopped with 15  $\mu$ L stop solution (90% formamide, 20 mmol/L EDTA, 0.01% bromphenol blue, and 10% ficoll), and the PCR products were denatured by heating at 95 °C for 2 to 4 minutes and then immediately incubating on ice for 5 minutes.

Denatured PCR products were run on 8% polyacrylamide gels (1.5 mm thickness x 28 cm length) at 60 w for 5 hours, after the gel was prerun at 70 w for 45 minutes. Gel solution was prepared with 8% acrylamide/bis-acrylamide (29/1), 30% formamide, 1x TAE, and 5.6 mol/L urea. TEMED (1%) and ammonium persulphate (0.08%) were added into the gel solution just before it was poured. After electrophoresis, the gel was transferred onto a UV transparent Plexiglass plate, stained with 2x SYBR-gold stain solution (Molecular Probes, Eugene, Oregon) in the dark for at least 30 minutes, and observed and imaged under UV light.

### ***Gene mapping strategies***

To identify the molecular markers that are linked to the pale mutation, we used the near-isogenic lines (NILs) screening strategy (Young et al. 1988), and the bulked segregant analysis (BSA) strategy (Michelmore et al. 1991).

In the NILs screening experiment, the Harosoy and Harosoy *w4* isoline samples were screened for polymorphisms with SSR markers. In the BSA experiment, 2 bulks were constructed from the F<sub>2</sub> mapping population [Minsoy(purple flowers) x T369(pale flowers)]. Bulk 1 contained DNA from 10 F<sub>2</sub> plants homozygous for wild-type purple flowers; Bulk 2 contained DNA from 10 F<sub>2</sub> plants homozygous for recessive pale flowers. Homozygotes were identified by progeny tests in the F<sub>2.3</sub> generation.

### ***Linkage analysis***

The Linkage-1 program (Suiter, et al. 1983) was used to evaluate the linkage between each candidate SSR marker and the locus that conditions the pale mutation by computing the recombination value between the pair.

The final map was constructed with the Mapmaker 2.0 program (Lander et al. 1987). Two markers were considered to be linked if the LOD score between them was equal to or higher than 3.0 and the recombination value between them was equal to or lower than 0.4. The genetic distance between 2 markers was generated from recombination rates by using the Kosambi map function (Kosambi 1944).

## Results

### Genetic analyses

The results of progeny tests of crosses of Minsoy, Harosoy *w1*, and Harosoy *w4* with T369 in the F<sub>1</sub>, F<sub>2</sub>, and F<sub>3</sub> generations are shown in Tables 2 and 3.

F<sub>1</sub> plants from the cross of Minsoy (purple flowers) with T369 (pale flowers) were all wild-type purple flower plants (Table 2, Cross 1). In the F<sub>2</sub> generation, the segregation ratio was 181 purple to 69 pale, which fit a 3:1 F<sub>2</sub> phenotypic segregation ratio ( $\chi^2 = 0.90$ ; P = 0.34) (Table 2, Cross 1). In the F<sub>2:3</sub> families, 69 F<sub>2</sub> plants with pale flowers gave all pale-flower F<sub>3</sub> plants, 115 F<sub>2</sub> plants with purple flowers gave F<sub>3</sub> plants with purple flowers and with pale flowers, and 66 F<sub>2</sub> plants with purple flowers gave all purple-flower F<sub>3</sub> plants, which showed a 1: 2: 1 F<sub>2</sub> genotypic segregation ratio ( $\chi^2 = 1.67$ ; P = 0.43) (Table 2, Cross 1). These results showed that the pale mutation was conditioned by a single locus, and that it was recessive to the wild-type purple flower phenotype.

The cross of Harosoy *w4* (near-white flowers) with T369 (pale flowers) was made to determine the relationship between the *w4* locus and the locus that conditions the pale-flower phenotype (Table 2, Cross 2). All the F<sub>1</sub> plants from the cross were pale flowers. In the F<sub>2</sub>

generation, 1153 plants were pale flowers, and 397 plants were near-white flowers, which showed a 3:1 phenotypic segregation ratio ( $\chi^2 = 0.31$ ;  $P = 0.58$ ). Among these 1550  $F_2$  plants, 150 plants were randomly selected to advance to the  $F_{2:3}$  generation. As a result, 42  $F_{2:3}$  families were true breeding for pale flowers, 74  $F_{2:3}$  families segregated for pale flowers and near-white flowers, and 34  $F_{2:3}$  families were true breeding for near-white flowers, which followed a 1: 2:1  $F_2$  genotypic segregation ratio ( $\chi^2 = 0.88$ ;  $P = 0.64$ ). The results (Table 2, Cross 2) suggested that the pale-flower mutation generated a new recessive allele ( $w4-p$ ) at the  $W4$  locus, and that it was dominant to the previously identified recessive  $w4$  allele.

The results of progeny tests for Harosoy  $w1$  (white flowers) x T369 (pale flowers) are shown in Table 3. The  $F_1$  plants were all purple-flower plants. The  $F_2$  segregation was 812 purple-flower plants: 245 pale-flower plants: 354 white-flower plants, and showed a 9:3:4 phenotypic segregation ratio ( $\chi^2 = 1.87$ ;  $P = 0.39$ ) (Table 3). The 165  $F_{2:3}$  families were classified into 7 categories: **1)** in families producing all purple flowers,  $F_2$  plants were genotype  $W1W1W4W4$ ; **2)** in families producing purple and pale flowers,  $F_2$  plants were genotype  $W1W1W4w4-p$ ; **3)** in families producing purple and white flowers,  $F_2$  plants were genotype  $W1w1W4W4$ ; **4)** in families producing purple, white, and pale flowers,  $F_2$  plants were genotype  $W1w1W4w4-p$ ; **5)** in families producing all pale flowers,  $F_2$  plants were genotype  $W1W1w4-pw4-p$ ; **6)** in families producing pale and white flowers,  $F_2$  plants were genotype  $W1w1w4-pw4-p$ ; and **7)** in families producing all white flowers,  $F_2$  plants were genotype  $w1w1\_ \_$ . The ratio followed a 1:2:2:4:1:2:4 pattern ( $\chi^2 = 4.24$ ;  $P = 0.64$ ) (Table 3). The results showed that 2 unlinked loci,  $w4-p$  and  $w1$ , determined the flower color of the plants in this cross, and that the  $w1$  locus had an epistatic effect on the  $w4-p$  mutation. The

plants with *w1w1* genotype produced white flowers irrespective of the genotype at the *w4-p* locus.

### **Molecular mapping of the *W4* gene**

An F<sub>2</sub> population with 114 individuals derived from a single F<sub>1</sub> plant from a cross between Minsoy (purple flowers, *W1W1W4W4*) and T369 (pale flowers, *W1W1w4-pw4-p*) was used to map the *W4* locus. The genotype of each F<sub>2</sub> plant was determined through a progeny test in the F<sub>2:3</sub> generation. The genotypic ratio of F<sub>2</sub> plants was 30 *W4W4*: 51 *W4w4-p*: 33 *w4-pw4-p*, and fit a 1:2:1 ratio ( $\chi^2 = 1.42$ ; P = 0.49) (Table 4).

A total of 178 SSR markers, representing all 20 MLGs of the soybean genome (Cregan et al. 1999), were selected to identify markers that were linked with the *W4* locus. The average genetic distances between any 2 adjacent markers were approximately 20 cM.

Two NILs, Harosoy (*W1W1W4W4*) and Harosoy *w4* (*W1W1w4w4*) were screened with the 178 SSR markers. Seven SSR markers, Satt430 on MLG B1, Sat\_130 on MLG C2, Satt548 on MLG D1a+Q, Satt372, Satt386, and Satt458 on MLG D2, and Satt288 on MLG G, were able to detect polymorphism between these 2 NILs on an 8% PAGE gel.

These 7 SSR markers and their flanking SSR markers were also used in the BSA experiment to screen the 2 bulks, Bulk 1 and Bulk 2, made from the F<sub>2</sub> mapping population, along with the 2 parental lines, Minsoy and T369. Bulk 1 contained 10 individuals homozygous for purple flowers, and Bulk 2 contained 10 individuals homozygous for pale flowers. As a result, Satt386 on MLG D2 was polymorphic between the 2 bulks and the 2 parents. The band amplified from Bulk 1 (purple flowers) was identical to the one from Minsoy (purple flowers), and the band amplified from Bulk 2 (pale flowers) was identical to

the one from T369 (pale flowers). Because in BSA strategy polymorphic molecular markers between 2 bulks should represent a DNA sequence within or adjacent to the selected gene (Michelmore et al. 1991), the SSR marker Satt386 was possibly closely linked to the *W4* locus.

To confirm this, Satt386 was used to screen the 114  $F_2$  plants of the mapping population. The segregation ratio of Satt386 was 30 AA (Minsoy pattern): 51 AB (heterozygote pattern): 33 BB (T369 pattern), which fit a ratio of 1: 2: 1 ( $\chi^2 = 0.58$ ;  $P = 0.75$ ) (Table 4). The recombination value between Satt386 and the *W4* locus was calculated as  $0.036 \pm 0.013$  by the Linkage-1 program (Suiter, et al. 1983), which showed that the *W4* locus was closely linked to the SSR marker Satt386 on soybean MLG D2.

To construct a high-resolution map for the *W4* locus, an additional 18 SSR markers from MLG D2 that were at a genetic distance of less than 65 cM from Satt386, were tested for polymorphism between the 2 parental lines of the mapping population, Minsoy and T369 (Fig. 2). Eleven markers were polymorphic between these 2 parental lines (Fig. 2). Five markers, Satt389, Satt226, Satt574, Satt186, and Satt413, which could clearly identify the polymorphism between the 2 parental lines on a 2% Agarose gel, were used to screen the  $F_2$  mapping population. The segregation ratio of each marker followed the ratio of 1AA (Minsoy pattern):2 AB (heterozygote pattern):1BB (T369 pattern) (Table 4).

Mapmaker 2.0 was used to generate a map for the *W4* locus, using the data from these 5 markers, along with the data from the *W4* locus and Satt386. The *W4* locus was mapped between Satt386 and the nearby telomere with the distance of 2.3 cM between *W4* and Satt386 (Fig. 2).

## Discussion

The probability of recovering new mutants is maximized by searching among progenies of germinal revertant (wild-type) plants descended from mutable plants. If the reversion of the unstable allele was the result of an excision of the element from its locus, new mutations might be detected among the progenies of these revertants. Such mutants would be expected if the excised element was inserted at a second locus.

The pale mutation was discovered among germinal revertant plants of the *w4*-mutable line. Genetic studies established that it was controlled by a new allele (*w4-p*) at the *W4* locus. The *w4-p* allele is dominant to the *w4* allele, but recessive to the *W4* allele. The *w1* locus has an epistatic effect on the *w4-p* allele. The mutants with the *w1w1* genotype would produce white flowers regardless of the genotype at the *W4* locus.

Two mutations for flowers with intermediate color were found among self-pollinated progenies of the *w4*-mutable lines. They are “dilute purple” (T321, *w4-dp*) (Palmer and Goose 1993) and “pale” (T369, *w4-p*). Both of them were identified as alleles at the *W4* locus. The molecular mechanisms are not known for these 2 alleles that evolved from the *w4-m* allele. Studies show that genes with a transposon insertion are hotspots for secondary mutations due to imprecise excision of the transposon, such as a few base changes in host sequence, and the host sequence duplication or deletion. The transposon also may jump into nearby regions and transpose back to the target gene to generate new types of mutant alleles for the target gene (Walbot 2000). Thus, it is possible that the *w4-p* and *w4-dp* alleles were generated through an imprecise excision of the transposable element harbored in the *w4-m* allele, or generated by a re-insertion of the transposon that inserted adjacent to the *w4-m* allele. Molecular studies are needed to address this question.

In the molecular mapping study, the *W4* locus was mapped to the terminus of MLG D2 between Satt386 and the nearby telomere with the distance of 2.3 cM between *W4* and Satt386 (Fig. 2). Compared with the USDA/Iowa State University molecular map (Cregan, et al. 1999), SSR markers including Satt389, Satt226, Satt574, Satt186, Satt413, and Satt386 followed the same order in our map, but the genetic distances between 2 adjacent markers were shorter most of the time than the genetic distances between the same 2 markers in the USDA/Iowa State University map (Fig. 2). One reason for this may be that the parental lines of the mapping population we used were different from the parental lines used to construct the USDA/Iowa State University map. Another reason may be that the same chromosomal region is more saturated with markers in the USDA/Iowa State University map. Therefore, in the same chromosome region, the undetectable even-numbered crossovers (mainly double crossovers) in our map would be recovered in the USDA/Iowa State University map, which would lengthen their map.

The results we obtained from molecular mapping of the *W4* locus are in agreement with the hypothesis that an active transposable element resides in the *w4*-mutable line. The *W4* locus was mapped to MLG D2; however, mutants found in independent germinal revertants of the *w4*-mutable line were mapped at different positions on different chromosomes. For example, mutant female partial-sterile 1 (*Fsp2*), female partial-sterile 2 (*Fsp3*), female partial-sterile 3 (*Fsp4*), and female partial-sterile 4 (*Fsp5*) were located on MLG C2, A2, F, and G, respectively (Kato and Palmer 2004). Mutant male-sterile and female-sterile (*st8*) was positioned on MLG J (Kato and Palmer 2003). Mutant chlorophyll-deficient leaves (CD-5) was located on MLG E (K.K. Kato and R.G. Palmer, unpublished data). And, mutant chlorophyll-deficient leaves (*y20*) and mutant malate dehydrogenase 1



null (*mdh1-n*) were closely linked on MLG H (unpublished data). It would be difficult to explain how the *w4-m* allele affects so many different genes *in trans* if there was no active transposable element in the *w4*-mutable line.

Transposon mutagenesis using active transposable elements is a very useful tool in gene cloning and functional genomics research in plants (Maes, et al. 1999; Walbot 2000; Ramachandran and Sundaresan 2001); however, no active transposable element has ever been found in soybean. Therefore, identifying and cloning of an active transposon in soybean is very important. A good way to do this is to try to use the mutable alleles found in soybean. The mutable alleles are the genes that most likely harbor an autonomous transposable element.

Besides the *w4-m* allele, there are 3 other mutable alleles reported in soybean that may contain an autonomous transposable element, the *Y18-m* allele (Peterson and Weber 1969), the *r-m* allele (Chandlee and Vodkin 1989a), and the *wp-m* allele (Johnson et al. 1998). The mutable line with the *Y18-m* allele displays variegated green/yellow leaves, the mutable line with the *r-m* allele produces seeds with variegated seed coat, and the mutable line with the *wp-m* allele produces variegated flower color similar to the *w4*-mutable line.

If these 4 mutable lines do harbor an autonomous transposable element, the *w4*-mutable line and the *wp*-mutable line have more advantages than the other 2 in transposon tagging experiments. First, their germinal revertant plants are more vigorous and healthier than most of the revertants of the *Y18-m* mutable line. The *Y18-m* allele usually generates the recessive *y18* allele that produces lethal-yellow plants (Chandlee and Vodkin 1989b). Moreover, some of the transposition-induced new mutations, such as chlorophyll deficiency, may not be readily identified in the revertants among the *Y18-m* mutable line. Second, flower

color is less affected by environments than seed-coat color; thus, the germinal revertants of the *w4-m* and *wp-m* mutable lines are easier to identify correctly than those of the *r-m* mutable line.

Finally, to clone the *w4-m* gene and the putative transposable element in the *w4-m* gene through chromosome walking, a higher-resolution map and a larger mapping population are needed. In conclusion, molecular mapping of the *W4* locus is the first step for positioned cloning of this gene, and the data provided more evidence to support the hypothesis that there is an autonomous transposable element in the *w4-m* allele.

## Acknowledgements

The authors would like to thank Dr. R.W. Groose, Department of Plant, Soil and Insect Sciences, University of Wyoming, Laramie, Wyo., for providing flower pictures. The authors would also like to thank Dr. P. B. Cregan, USDA-ARS, Soybean and Alfalfa Research Lab., Beltsville, Md., for providing the protocols for PCR and PAGE in the SSR experiments.

## References

- Akkaya, M.S., Bhagwat, A.A., and Cregan, P.B. 1992. Length polymorphism of simple sequence repeat DNA in soybean. *Genetics*, **132**: 1131-1139.
- Akkaya, M.S., Shoemaker, R.C., Specht, J. E., Bhagwat, A. A., and Cregan, P. B. 1995. Integration of simple sequence repeat DNA markers into a soybean linkage map. *Crop Sci.* **35**: 1439-1445.

- Chandlee, J.M., and Vodkin, L.O. 1989a. Unstable expression of a soybean gene during seed coat development. *Theor. Appl. Genet.* **77**: 587-594.
- Chandlee, J.M., and Vodkin, L.O. 1989b. Unstable genes affecting chloroplast development in soybean. *Dev. Genet.* **10**: 532-541.
- Cregan, P.B., Bhagwat, A.A., Akkaya, M.S., and Jiang, R. 1994. Microsatellite fingerprinting and mapping of soybean. *Meth. Mol. Cell. Biol.* **5**: 49-61.
- Cregan, P.B., Jarvik, T., Bush, A.L., Shoemaker, R.C., Lark, K.G., Kahler, A.L., Kaya, N., VanToai, T.T., Lohnes, D.G., Chung, J., and Specht, J.E. 1999. An integrated genetic linkage map of the soybean genome. *Crop Sci.* **39**: 1464-1496.
- Groose, R.W., and Palmer, R.G. 1991. Gene action governing anthocyanin pigmentation in soybean. *J. Hered.* **82**: 498-501.
- Groose, R.W., Schulte, S.M., and Palmer, R.G. 1990. Germinal reversion of an unstable mutation for anthocyanin pigmentation in soybean. *Theor. Appl. Genet.* **79**: 161-167.
- Groose, R.W., Weigelt, H.D., and Palmer, R.G. 1988. Somatic analysis of an unstable mutation for anthocyanin pigmentation in soybean. *J. Hered.* **79**: 263-267.
- Hartwig, E.E., and Hinson, K. 1962. Inheritance of flower color in soybeans. *Crop Sci.* **2**: 152-153.
- Hegstad, J.M., Tarter, J.A., Vodkin, L.O., and Nickell, C.D. 2000. Positioning the *wp* flower color locus on the soybean genome map. *Crop Sci.* **40**: 534-537.
- Johnson, E.O.C., Stephens, P.A., Fasoula, D.A., Nickell, C.D., and Vodkin, L.O. 1998. Instability of a novel multicolored flower trait in inbred and outcrossed soybean lines. *J. Hered.* **89**: 508-515.

- Kato, K.K., and Palmer, R.G. 2003. Molecular mapping of a male-sterile, female-sterile mutant gene (*st8*) in soybean. *J. Hered.* **94**: 425-428.
- Kato, K.K., and Palmer, R.G. 2004. Molecular mapping of four ovule lethal mutants in soybean. *Theor. Appl. Genet.* **108**: 577-585.
- Keim, P., Olson, T., and Shoemaker, R.C. 1988. A rapid protocol for isolating soybean DNA. *Soybean Genet. Newsl.* **15**: 150-152.
- Kosambi D.D. 1944. The estimation of map distance from recombination values. *Ann. Eugen.* **12**: 172-175.
- Lander, E.S., Green, P., Abrahamson, J., Barlow, A., Daly, M.J., Lincoln, S.E., and Newburg, L. 1987. MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics*, **1**: 174-181.
- Maes, T., De Keukeleire, P., and Gerats, T. 1999. Plant tagnology. *Trends Plant Sci.* **4**: 90-96.
- Michelmore, R.W., Paran, I., and Kesseli, R.V. 1991. Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proc. Natl. Acad. Sci. USA*, **88**: 9828-9832.
- Palmer, R.G., Groose, R.W., Weigelt, H.D., and Miller, J.E. 1990. Registration of a genetic stock (*w4-mw4-m*) for unstable anthocyanin pigmentation in soybean. *Crop Sci.* **30**: 1376-1377.
- Palmer, R.G., and Groose, R.W. 1993. A new allele at the *w4* locus derived from the *w4-m* mutable allele in soybean. *J. Hered.* **84**: 297-300.

- Palmer, R.G., Hedges, B.R., Benavente, R.S., and Groose, R.W. 1989. *w4*-mutable line in soybean. *Dev. Genet.* **10**: 542-551.
- Palmer, R.G., Pfeiffer, T.W., Buss, G.R., and Kilen, T.C. 2004. Qualitative genetics. In: Specht, J.E. and Boerna H.R. (eds) *Soybeans: Improvement, production, and uses*, 3<sup>rd</sup> ed. Am. Soc. Agron. Madison, WI, pp. 137-234.
- Palmer, R.G., Harper, T.K., and Martin, S.K. 1997. A mutable flower phenotype from Asgrow XP2214. *Soybean Genet. Newsl.* **24**: 176-178.
- Peterson, P.A., and Weber, C.R. 1969. An unstable locus in soybeans. *Theor. Appl. Genet.* **39**: 156-162.
- Ramachandran, S., and Sundaresan, V. 2001. Transposons as tools for functional genomics. *Plant Physiol. Biochem.* **39**: 243-252.
- Stephens, P.A. and Nickell, C.D. 1992. Inheritance of pink flower color in soybean. *Crop Sci.* **32**: 1131-1132.
- Suiter, K.A., Wendel, J.F., and Case, J.S. 1983. Linkage-1: a PASCAL computer program for the detection and analysis of genetic linkage. *J. Hered.* **74**: 203-204.
- Walbot, V. 2000. Saturation mutagenesis using maize transposons. *Curr. Opin. Plant Biol.* **3**: 103-107.
- Young, N.D., Zamir, D., Ganai, M.W., and Tanksley, S.D. 1988. Use of isogenic lines and simultaneous probing to identify DNA markers tightly linked to the *Tm-2a* gene in tomato. *Genetics*, **120**: 579-585.

Table 1. Genotypes and phenotypes of soybean lines used in the genetic and molecular mapping experiments.

Lines	Genotype	Description	Flower color
Harosoy	<i>W1W1W4W4</i>	True breeding wild-type soybean	Purple
Harosoy <i>w4</i> (L72-1138)	<i>W1W1w4w4</i>	True breeding stable recessive <i>w4</i> mutant	Near white
Harosoy <i>w1</i> (L62-906)	<i>w1w1W4W4</i>	True breeding stable recessive <i>w1</i> mutant	White
T369	<i>W1W1w4-pw4-p</i>	True breeding stable revertant from the <i>w4</i> -mutable line	Pale
Minsoy (PI 27890)	<i>W1W1W4W4</i>	True breeding wild-type soybean	Purple

Table 2. Progeny tests of crosses of Minsoy (purple flowers) and Harosoy *w4* (L72-1138) (near-white flowers) with T369 (pale flowers) in the F<sub>1</sub>, F<sub>2</sub>, and F<sub>3</sub> generations.

Cross*	F <sub>1</sub> phenotype	Segregation in the F <sub>2</sub> generation					Segregation in the F <sub>3</sub> generation						
		No. of plants			$\chi^2_{(3:1)}$	P	No. of F <sub>2:3</sub> families					$\chi^2_{(1:2:1)}$	P
		Pu	pa	wh			Pu	Pu/pa	pa	pa/wh	wh		
1	All purple	181	69		0.9	0.34	66	115	69			1.67	0.43
2	All pale		1153	397	0.31	0.58			42	74	34	0.88	0.64

\***Cross 1:** Minsoy (purple flowers) x T369 (pale flowers); **Cross 2:** Harosoy *w4* (near-white flowers) x T369 (pale flowers)

**Note:** **Pu**, plants or families that produced all purple flowers; **pa**, plants or families that produced all pale flowers; **wh**, plants or families that produced all near-white flowers; **Pu/pa**, families consisting of purple-flower plants and pale-flower plants; **pa/wh**, families consisting of pale-flower plants and near-white-flower plants.

Table 3. Progeny tests of the cross Harosoy *w1* (L62-906) (white flowers) with T369 (pale flowers) in the F<sub>2</sub> and F<sub>2:3</sub> generations.

	Segregation in the F <sub>2</sub> generation			Family segregation in the F <sub>2:3</sub> generation						
	Pu	pa	wh	Pu	Pu/pa	Pu/wh	Pu/pa/wh	pa	pa/wh	wh
No. plants/families	812	245	354	11	17	19	52	9	18	39
Genotype of plants/families	<i>W1-</i> <i>W4-</i>	<i>W1-</i> <i>w4-pw4-p</i>	<i>w1w1</i> --	<i>W1W1</i> <i>W4W4</i>	<i>W1W1</i> <i>W4w4-p</i>	<i>W1w1</i> <i>W4W4</i>	<i>W1w1</i> <i>W4w4-p</i>	<i>W1W1</i> <i>w4-pw4-p</i>	<i>W1w1</i> <i>w4-pw4-p</i>	<i>w1w1</i> --
Expected no.	794	265	353	10	21	21	41	10	21	41
$\chi^2_{(9:3:4)} = 1.87$ P = 0.39			$\chi^2_{(1:2:2:4:1:2:4)} = 4.24$ P = 0.64							

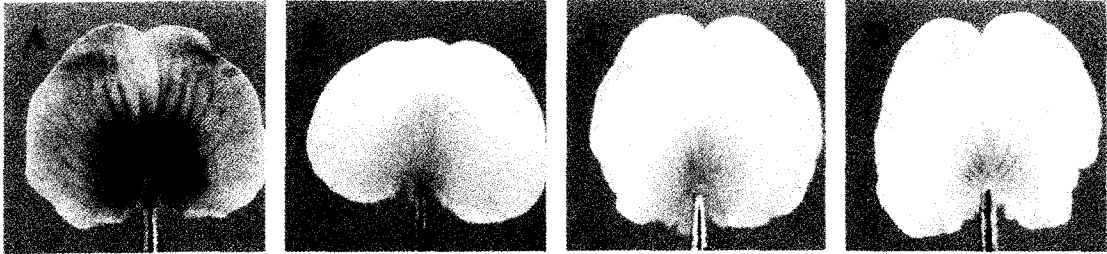
**Note:** **Pu/wh**, families consisting of purple-flower plants and white-flower plants; **Pu/pa/wh**, families consisting of purple-flower plants, pale-flower plants, and white-flower plants.



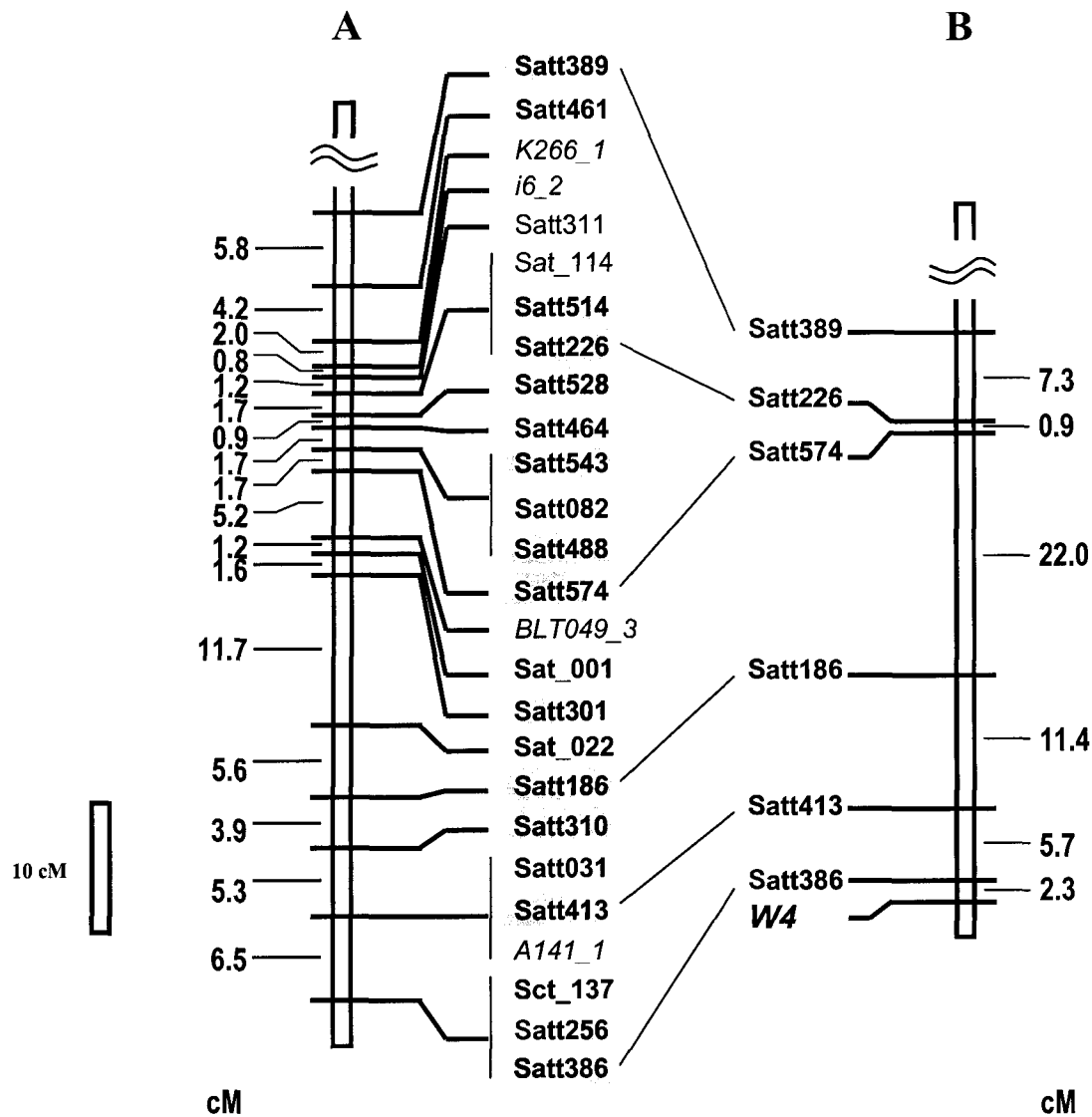
Table 4. SSR marker segregation patterns in the mapping population of Minsoy (purple flowers) x T369 (pale flowers).

Markers	Number of F <sub>2</sub> plants			df	$\chi^2_{(1:2:1)}$	P
	AA	AB	BB			
<i>W4</i>	30	51	33	2	1.42	0.49
Satt386	30	53	31	2	0.58	0.75
Satt413	31	57	26	2	0.44	0.8
Satt186	31	59	24	2	1.00	0.61
Satt574	32	60	22	2	2.07	0.36
Satt226	32	60	22	2	2.07	0.36
Satt389	36	57	21	2	3.95	0.14
Expected no.	28.5	57	28.5			

**Note:** Genotypes of F<sub>2</sub> plants: **AA**, homozygous Minsoy genotype; **AB**, heterozygous genotype; **BB**, homozygous T369 genotype.



**Fig. 1** Standard petals of flowers of soybean wild-type plants and mutants. (A) wild-type purple Harosoy (*W1W1W4W4*); (B) near-white Harosoy *w4* isoline (L72-1138) (*W1W1w4w4*); (C) and (D) pale-mutant line, T369, (*W1W1w4-pw4-p*).



**Fig. 2** Comparison of the molecular linkage map of MLG D2 (USDA/ Iowa State University) and the molecular linkage map of the *W4* locus. Distances are shown in centiMorgans (cM). (A) MLG D2 from USDA/ Iowa State University map (Cregan et al. 1999). Markers in bold are those screened for the two parental lines of the mapping population, Minsoy and T369. Markers highlighted in grey are the ones that detected polymorphisms between Minsoy and T369. (B) Molecular linkage map of the *W4* locus constructed in this study.

**CHAPTER 3 THE *W4* LOCUS CORRESPONDS TO A  
DIHYDROFLAVONOL-4-REDUCTASE (DFR) GENE IN SOYBEAN  
[*GLYCINE MAX* (L.) MERR.]**

Min Xu and Reid G. Palmer

**Abstract**

In soybean, the *W4* locus is one of the loci that control anthocyanin biosynthesis of soybean flowers and hypocotyls. A putative transposable element was suggested to reside within or adjacent to it in the mutable line T322 (*w4-m*). In present study, the immature flower petals of six samples from five soybean lines with different *W4* alleles including Harosoy (*W4*), Harosoy *w4* (*w4*), T321 (*w4-dp*), T322 (*w4-m*), and T369 (*w4-p*) were used for spectrophotometry, HPLC, RT-PCR, and Northern blot analyses. The white petal sectors in T322 (T322w) and the purple petal sectors of T322 (T322p) were collected and examined separately. Results showed delphinidin and its derivatives were the main pigments in soybean flowers. Their contents were lower in flowers with less pigmentation. This was associated with low transcript level or abnormal transcript products of the DFR2 gene. RFLP analysis, conducted with a F<sub>2</sub> mapping population of Minsoy X T369 (*w4-p*), showed that the *W4* locus co-segregated with a DFR gene, indicating that the *W4* locus either contained a DFR gene or was close to one. According to the results from spectrophotometry, HPLC, RT-PCR, and Northern blot analyses, the *W4* locus very likely corresponds to the DFR2 gene.

## Introduction

The *W4* locus is a major locus that controls anthocyanin biosynthesis of flowers and hypocotyls in soybean. Mutations at this locus would reduce the purple pigment in flowers and hypocotyls. A total of four mutant alleles were found at this locus. The *w4* allele gives soybean with near-white flowers and green hypocotyls (Hartwig and Hinson 1962; Groose and Palmer, 1991). The *w4-dp* allele leads to dilute purple flowers and green hypocotyls (Palmer and Groose, 1993). The *w4-p* allele leads to pale flowers and green hypocotyls (Xu and Palmer, 2005). The *w4-m* allele gives variegated flowers and hypocotyls (Groose et al., 1988). The color of the standard petal for each allele is shown in Fig. 1.

The *w4-m* allele was proposed to contain an autonomous transposable element (Palmer et al., 1989). Many mutants were found among the descendents of the germinal revertants of the mutable lines. The genetic stock with homozygous *w4-m* alleles was registered as the *w4*-mutable line, and the Genetic Type Collection number for this line is T322 (Palmer et al., 1990). The *w4-dp* and *w4-p* alleles, identified in gene tagging experiments, were possibly generated through partial reversion. By using SSR markers, the *W4* locus was located at the end of molecular linkage group D2 (Xu and Palmer, 2005), which is at a different location from other mutants found in the gene tagging experiment with T322. For example, mutant male-sterile and female-sterile (*st8*) was positioned on MLG J (Kato and Palmer, 2003). Mutants female partial-sterile 1 (*Fsp2*), female partial-sterile 2 (*Fsp3*), female partial-sterile 3 (*Fsp4*), and female partial-sterile 4 (*Fsp5*) were located on MLG C2, A2, F and G, respectively (Kato and Palmer, 2004). And, mutant chlorophyll-deficient leaves (*y20*) and mutant malate dihydrogenase1 null (*Mdh1-n*) were tightly linked on MLG H (Xu and Palmer, unpublished data).

Transposable elements (TEs) including endogenous TEs and heterologous TEs are very useful tools in gene cloning and functional genomics (Ramachandran and Sundaresan, 2001; Walbot, 2000; Maes et al., 1999). In soybean, the production of transgenic soybean plants remains inefficient although efforts are being made to improve the transformation efficiency (Trick et al., 1997). Heterologous TEs from other species are difficult to use in soybean. Thus, identifying and cloning active transposons from a soybean mutable line such as T322 would be very useful for gene cloning and functional genomics projects.

The mutable allele in T322 was found at the *W4* locus. This locus affects anthocyanin biosynthesis of flowers and hypocotyls. Anthocyanin biosynthesis pathway is well characterized in plants (Fig. 2) (Holton and Cornish, 1995). It is conditioned by two types of genes, structural genes that encode anthocyanin biosynthetic enzymes [eg. CHS (chalcone synthase), F3H (flavanone 3-hydroxylase), DFR, ANS (anthocyanin synthase) *etc*], and regulatory genes that control the expression of structural genes.

In this project, the objective was to identify the role of the *W4* gene in soybean anthocyanin biosynthesis pathway through both biochemical and molecular biological methods.

## **Materials and methods**

### **Plant materials**

Five soybean lines with different *W4* alleles (Harosoy, Harosoy *w4*, T321, T322, and T369) were planted in the Bruner Farm (Ames, IA). Their genotype and flower color were described in Fig. 1. Flower petals were collected from floral buds 1 day before blooming for

analyses of anthocyanin, flavonol, and RNA. Here, the purple petal sections of T322 (T322p) were collected separately from the white petal sections (T322w).

A F<sub>2</sub> mapping population derived from cultivar Minsoy X T369 (Xu and Palmer, 2005) was used for restriction fragment length polymorphism (RFLP) analysis. This population was used to map the *W4* locus in soybean.

### **Extraction and analysis of anthocyanidins**

Flower-petal samples were collected, freeze-dried, and ground. The anthocyanin pigments were extracted in 1% (v/v) HCL in methanol for 3 hours at room temperature. The extracts were centrifuged at 13,000 rpm for 10 minutes. Half of the supernatants were used for spectrophotometry analysis with a Beckman DU 640 nucleic acid and protein analyzer. The other half was hydrolyzed by boiling for 30 minutes. Hydrolyzed extracts also were subjected to spectrophotometry analysis. The anthocyanidin contents were expressed as the absorbance at 535nm ( $A_{535}$ ) per mg dried petals per mL solvent.

### **High performance liquid chromatography (HPLC) analysis of flavonols**

The flavonol aglycones of soybean flowers and authentic standard solutions of myricetin, quercetin, and kaempferol (Sigma, St. Louis, MO.) were prepared according to the method previously used in *Arabidopsis* (Burbulls et al., 1996). Freeze-dried flower petals (50 mg in fresh weight) were ground to fine powders, suspended in 500  $\mu$ L of 80% (v/v) methanol, and incubated at room temperature for 10 min. The mixtures were centrifuged at 15,000 g for 5 min. The supernatants were hydrolyzed by incubating in 1N HCl at 70°C for 40 min. Then, an equal volume of 100% methanol was added into the hydrolyzed extracts to prevent precipitation of flavonol aglycones.

The flavonol aglycones were separated using a modified method of Burbulls et al. (1996). The samples (100 µL) were injected onto a RP-18 column and eluted at a flow rate of 1.0 mL/min with a linear gradient of HPLC-grade acetonitrile in HPLC-grade H<sub>2</sub>O (pH 3.0, adjusted with glacial acetic acid) as following: 0 to 0% for 5 min, 0 to 10% for 5 min, 10 to 30 % for 60 min, 30 to 100% for 5 min, 100 to 100% for 2 min, 100 to 0%, for 2 min, and 0-0% for 5 min. The elutant was detected at 255 nm with photo-diode array 996(PDA996), and quantified by comparing to the standards.

### **DNA and RNA preparation**

Genomic DNA was extracted from young leaves with CTAB method (Keim et al., 1988). Total RNA was prepared from immature flowers of five soybean lines using Trizol reagent (Invitrogen, Carlsbad, CA) according to the manufacturer's protocol.

### **Reverse transcriptase (RT)-PCR**

Total RNA (1 µg) from each flower sample was used as the template. The RT-PCR reactions were conducted with Onestep RT-PCR kit (Qiagen, Valencia, CA) by using the manufacturer's protocol. The primers for synthesizing the cDNA of DFR gene were DFR-f (5'- ATCTTGCTGAAGAGGGAAGC -3') and DFR-r (5'- TGTACATGTCCTCTAAGCTG -3'). This pair of primers was designed according to the consensus sequence of three DFR genes from three legumes, *Glycine max* [gi:5852932], *Lotus corniculatus* [gi:31324463], and *Medicago truncatula* [gi: 38683950].

The primers for synthesizing the cDNA for ANS gene were ANS-f (5'- ATGCACTTGGTGAACCATGG -3') and ANS-r (5'- ATGCACTTGGTGAACCATGG -3'). They were designed from a partial coding sequence of anthocyanin synthase mRNA



identified in seed coat [gi: 14028670]. By blasting to the expressed sequence tag (EST) data base, this *ANS* gene was expressed in many tissues such as seed coat, mature and immature flowers, and young seedlings.

The annealing temperature for amplifying both genes was 54 °C.

### **Northern blot analysis**

Twenty microgram of total RNA per sample was separated on a 1.0% formaldehyde-agarose gel (Thomas, 1980) and blotted onto a Zeta Probe Nylon membrane (BioRad, Hercules, CA) by capillary transfer. <sup>32</sup>P-labeled cDNA probes for DFR, F3H, and ANS genes were prepared with Primer-it II randomly labeling kit (Stratagene, La Jolla, CA). The probes for DFR and ANS genes were purified from RT-PCR products of cv. Harosoy amplified as described in above section. The probe for F3H gene was prepared from an EST clone [gi:17022852] provided by R. C. Shoemaker.

### **RFLP analysis**

Genomic DNAs (5 µg) of F<sub>2</sub> individuals and their parental lines of the *W4* mapping population (Xu and Palmer, 2005), made from Minsoy X T369, were digested with restriction enzyme(s) at 37°C overnight and separated on a 0.8% (w/v) agarose gel. The Southern blot analysis was conducted as previously described and detected with <sup>32</sup>P-labeled probes (Sambrook et al., 1989).

## Results

### **Delphinidin or its derivatives are the major pigment in soybean flowers.**

Three types of anthocyanin (red cyanidin, brick-red pelargonidin, and blue delphinidin) are synthesized in plant. These anthocyanidins may be modified further by glycosylation, methylation, and acylation in many species (Holton and Cornish, 1995). In soybean, the main pigment of hypocotyl, stem, and subepidermal is malvidin (Nozzolillo, 1973), which is a derivative from delphinidin after glycosylation and methylation.

The methanol-HCl extracts of flower petals from Harosoy had a large spectrum absorption peak at 535nm, and the absorption peak shifted about 8nm to 543nm after hydrolyzation (Fig. 3). The methanol-HCl extracts from the other five samples (Harosoy *w4*, T322w, T322p, T321, and T369) showed similar spectrum characteristics (data not shown). These spectrum characteristics were the same as delphinidin-3-monoglucocide and its derivatives, petunidin-3-monoglucocide and malvidin-3-monoglucocide (Harborne, 1958). It indicated that the main pigments in flowers are delphinidin and its derivatives, same as in hypocotyls and stems.

Among six flower samples, anthocyanins were synthesized most in wild-type purple flowers [Harosoy and T322(p)], second in pale flowers (T369), third in dilute purple flowers (T321), and least in near-white flowers (Harosoy *w4*) and white sectors of T322 (T322w) (Fig. 4A), which were consistent with their phenotypes (Fig. 1).

### **The myricetin contents accumulated in flower petals with less pigment**

In plant, dihydroflavonols are the branch point of synthesis of flavonols and anthocyanins. It can be converted to leucoflavonols by DFR, which is the precursor of

anthocyanin, or flavonols by flavonol synthase (FLS) (Strack, 1997). The flavonol contents are supposed to be positively correlated with the dihydroflavonols contents if FLS functions normally. In this experiment, we checked the levels of the myricetin contents in six flower-petal materials from five investigated soybean lines with different *W4* alleles (Harosoy, Harosoy *w4*, T322p, T322w, T321, and T369) with HPLC. Myricetin is the product of dihydromyricetin by FLS, and dihydromyricetin is the precursor of delphinidin pigments, which were suggested to be the main pigments of soybean flowers.

HPLC analysis showed that the myricetin contents were higher in the mutant flowers with less anthocyanin pigments such as Harosoy *w4*, T321 (*w4-dp*), T322w (*w4-m*), and T369 (*w4-p*) (Fig. 4B). Except in flowers of T369, the anthocyanin contents were basically negatively correlated with the myricetin contents. In T369, the myricetin contents were extremely high, and its anthocyanin contents were also higher than all the other mutants at the *W4* locus.

#### **The mainly expressed DFR gene in immature soybean flowers was not the DFR1 gene [gi:5852932]**

Through RT-PCR method, a 743 bp cDNA fragment was amplified using DFR primers from all the samples except T322w, and a 621 bp cDNA fragment was amplified using ANS primers from all six samples (Fig. 4C). These two fragments amplified from Harosoy and T322p were sequenced with ABI 3730 DNA analyzer. Sequencing results showed that genes amplified from the immature flowers of Harosoy and T322p were same.

The ANS cDNA fragment amplified from soybean immature flowers was the same as the ANS gene identified in seed coat [gi:14028670]; however, the DFR cDNA fragments

amplified from immature flowers were different from the DFR1 gene [gi:5852932] identified in young seed coats. The homology between the DFR1 gene and the cDNA fragment amplified by DFR primers was 84%, and the homology between the deduced amino acid sequences of the two genes was 81%. We named the deduced protein dihydroflavonol-4-reductase 2 (DFR2). The partial DFR2 cDNA sequence and deduced amino acid sequence were submitted to Genbank and NCBI database. Its access number is DQ026299.

In Harosoy *w4*, another fragment with the size of ~1000bp was weakly amplified (Fig. 4C). This band could have come from an unspecific amplification, because it would be produced several weak bands after being re-amplified with the DFR primers (data not shown).

#### **The DFR2 gene had lower transcript level in mutant flowers**

We measured the mRNA levels of three structural genes in immature flower petals of six samples with Northern blot and/or RT-PCR analyses. These genes encode F3H, DFR, and ANS, respectively. The results showed that the transcript levels of F3H and ANS genes were similar among all the examined soybean lines (Fig. 4C and D). Northern blot analysis showed that transcript level of the ANS gene in wild-type purple flowers (Harosoy and T322p) was even slightly lower than in mutant flowers (Fig. 4D), which was probably caused by anthocyanins through end-product repression. End-product repression is an important regulatory mechanism in biosynthetic pathways in bacteria system (for review, see Bright et al. 1979). In plant, some biosynthetic pathways also were observed the phenomenon of end-product repression, for example, the GA (gibberellins) biosynthetic pathway in *Arabidopsis* (Xu et al. 1995).

The transcript levels of DFR genes in all six samples were examined using RT-PCR and Northern blot methods. Both experiment showed similar results; that is, the transcript amounts of DFR genes in flowers were positively correlated with their anthocyanin contents (Fig. 4C, D).

The amount of DFR2 cDNAs fragment amplified from T321, Harosoy *w4*, or T322w using RT-PCR was much less than that amplified from wild-type purple flowers Harosoy or T322p (Fig. 4C). The amplification from T322w was undetectable. This suggested that the transcripts of the DFR2 gene were synthesized less in T321, T322w, and Harosoy *w4*.

Northern blot analysis gave consistent results (Fig. 4D). The transcript abundance of DFR gene was 2-fold less in T369 and 8-fold less in T321 when compared to the wild-type Harosoy. In the mutable line T322, the transcript level of DFR gene in the purple sectors (T322p) was similar to Harosoy, but in the white sectors (T322w), the transcripts of DFR gene were undetectable. In mutant Harosoy *w4*, two mRNA bands were detected, but neither of them was the same size as the mRNA band in wild-type Harosoy or the other mutant lines. The upper band was larger than the bands amplified from other soybean lines, and was about 1-fold less in quantity than the band from Harosoy. The lower band was smaller than the band from other soybean lines, and was about 0.5-fold less in quantity than the band from Harosoy.

#### **The *W4* gene co-segregated with a DFR gene**

The F<sub>2</sub> individuals from a mapping population developed for the *W4* mapping experiment (Xu and Palmer, 2005) were used in the RFLP analysis. Minsoy and T369, which are parents of the mapping population, showed restriction fragment length polymorphism of

the DFR gene with *Eco*RI digestion. A ~ 5.0 kb band was detected in Minsoy but not in T369 while a ~6.5 kb band was detected in T369 but not in Minsoy (Fig. 5). This marker screened 36 individuals randomly chosen from the mapping population. Results showed that the DFR RFLP co-segregated with the *W4* locus (Fig. 5), which indicated that the *W4* gene was either a DFR gene or closely linked to one.

## Discussion

### **Lower expression of the DFR2 gene at transcriptional level was responsible for less pigment in flowers in soybean lines with mutations at the *W4* locus**

Spectrophotometry analysis showed that the main pigments in soybean flowers were delphinidin or its derivatives such as petunidin or malvidin. Their contents were lower in immature flowers of the soybean lines that mutated at the *W4* locus, such as Harosoy *w4*, T321 (*w4-dp*), T322 (*w4-m*), and T369 (*w4-p*) (Fig. 4A). However, their precursor, dihydromyricetin, the contents of which positively correlate to the contents of myricetin, was accumulated in these less pigmented flowers (Fig. 4B). These results indicated that mutation at the *W4* locus blocked the anthocyanin biosynthesis pathway after dihydromyricetin formed.

We checked the transcript level of three structural genes that encode F3H, DFR, and ANS (Fig. 4C, D). F3H catalyzes a reaction before the one catalyzed by DFR. ANS catalyzes a reaction after the one catalyzed by DFR (Fig. 2). Only transcripts from the DFR gene were observed to differ in quantity or molecular size among soybean lines with different *W4* alleles. The DFR transcripts had less amount in T321, T322w, and T369 and different sizes

in Harosoy *w4*. Through the RT-PCR results, the main expressed DFR gene in soybean immature flowers is *DFR2*.

Lower expression of the *DFR2* gene would lead to less production of *DFR2*, and set a blockage at the reduction step from dihydromyricetin to leucomyricetin, which *DFR* catalyzes. The blockage at this step would cause the substrate, dihydromyricetin, to accumulate, and indirectly result in higher myricetin contents, since the myricetin is the other product of dihydromyricetin. On the other hand, the blockage at this step could also cause less leucomyricetin to be produced, which further leads to less amount of delphinidin. This finally causes the phenotype of the less pigmented flowers.

#### **The *W4* locus may encode *DFR2* in soybean**

The transcripts of the *DFR2* gene had less quantity in the mutant flowers but the transcript levels of *F3H* and *ANS* genes were normal. So, mutation at the *W4* locus only had effects on the transcription of the *DFR2* gene, which implied that the *W4* locus encoded *DFR2*. This is because regulatory genes in anthocyanin biosynthesis pathway usually affect expressions of multiple structural genes (Holton and Cornish, 1995).

The results from RFLP analysis gave more support for this hypothesis. Through RFLP analysis, the *W4* locus was found to co-segregate with a *DFR* gene. This indicated that the *W4* locus was either the *DFR2* gene or a regulatory gene that was very close to a *DFR* gene and regulated the transcription of the *DFR2* gene, which is more unlikely to happen.

### **Alternate splicing could happened at the DFR2 gene in Harosoy *w4***

In Northern analysis, two DFR mRNA products were observed in immature flowers of Harosoy *w4*, but neither of them had the same size as the DFR2 mRNAs observed in other soybean lines.

From the results of spectrophotometry analysis and Northern blot, these two mRNA products should produce non-active or low-active DFR enzymes. These two analyses showed that the anthocyanin contents were positively correlated with the transcript level of DFR gene. In Harosoy *w4*, the amount of the upper mRNA band was about 1-fold less than Harosoy, and the amount of the lower mRNA band was about 0.5-fold less than Harosoy, so either band had higher density than the one in T321 or T369 which is 8-fold less or 2-fold less than Harosoy, respectively (Fig. 4D). Total DFR transcript levels in Harosoy *w4*, which combined these two bands, were even little higher than the DFR gene in wild-type flowers (Fig. 4D). However, the flowers of Harosoy *w4* had anthocyanin contents lower than the ones of T321, T369, and wild types, but similar to the ones of T322w whose DFR transcripts were undetectable (Fig. 4A). This means neither mRNA product in Harosoy *w4* could be translated to a DFR that functions well. These two mRNA products could result from expression of two different DFR genes or alternative splicing of the DFR2 gene.

Some of the structural genes in the anthocyanin pathway were clustered in a small chromosomal region. For example, in soybean, three CHS genes (CHS1, CH4, and CH5) were found in a 10 kb chromosomal region corresponding to the *I* locus on molecular linkage group A2 (Akada and Dube, 1995; Todd and Vodkin, 1996). These three genes were arranged as inverted repeats; the whole structure was duplicated at the adjacent chromosomal region; and together, these six genes formed a 27 kb chromosomal segment (Tuteja et al.,



2004). Normally, the *I* locus would silence the expression of CHS genes through posttranscriptional gene silencing, whereas deletion at the promoter of the *CHS4* gene or *ICHS1* could resume the transcript level of CHS genes. In rice, seven genes that encode DFR-like proteins were arranged in tandem within a 56 kb chromosome region on chromosome 4 (Lei et al., 2002).

Given that the two DFR mRNA products expressed in Harosoy *w4* were usually silenced by the intact *W4* locus that corresponded to the DFR2 gene, mutations at this locus would recover their expression. Then, according to the results from the present study, these two bands should show up in the flower sample of T322w, whose DFR2 gene is not detectable. However, no extra bands were observed in this sample. Therefore, this hypothesis may not be true.

If the two mRNA products detected by the DFR probe in Harosoy *w4* were not caused by two genes' expression, then alternative splicing would be another logical explanation. Alternative splicing happens at the post-transcription step of gene expression in eukaryotic genomes. It is a common event in mammals such as human and mouse. Through genome-wide survey, 40% alternative spliced transcripts were found in human genome, but in plant such as *Arabidopsis*, this number was much lower, just 5% (Brett et al., 2002). Alternative splicing could process in five different ways, splicing/not splicing, exon skipping, alternative 5' or 3' site, mutual exclusion of exons and retention of introns selection, and 3' exon choice (McKeown, 1992). Not only could alternative splicing produce structurally and/or functionally distinct proteins from a single gene and increase protein diversity in an organism, but also it could introduce mutations.

Likely, in Harosoy *w4*, the two mRNA products were generated from the DFR2 gene through alternative splicing, but neither is the right size. The inappropriate splicing could cause open reading frame (ORF) shifts, which would terminate translation prematurely and generate short non-functional amino acid fragments. If, on the other hand, the ORFs of the two mRNAs with inappropriate size were not shifted, they would produce an extended or truncated protein that would have a different structure from the normal DFR2. The alteration in structure would cause the enzyme to decrease or lose activity, which results in near-white flowers in Harosoy *w4*.

#### **DFR in soybean was encoded by multiple gene family**

In plant, many structural enzymes in anthocyanin biosynthesis pathway were encoded by multiple gene families. For example, CHS genes were reported to have at least eight isoforms in the soybean genome (Akada et al., 1990a, 1990b, 1991, 1993a, 1993b, 1993c; Akada and Dube, 1995).

DFR genes in the soybean genome also have many isoforms. Besides the *W4* locus, the *W3* locus was also found co-segregating with a DFR gene (Fasoula et al., 1995). Genetics analysis showed these two loci were unlinked (Palmer and Groose, 1993), which means they correspond to different DFR genes.

In NCBI database, including the DFR2 we submitted, a total of 34 mRNA sequences from soybean were deposited as homologues of DFR gene (Table 1). Three of them (gi:18691962, gi:11688949, and gi:5678147) are annotated incorrectly because they don't have homology with any of the DFR genes. The remaining 31 sequences from different

tissues of different soybean lines were classified into 7 groups representing 7 different *DFR* genes through BLAST, and we named them *DFR1-7* respectively in this paper.

Six of them (*DFR1-6*) were found only in certain tissues. These six genes could be tissue specific expressed genes; or their expression in other tissue was not identified. For example, the *W4* locus affects both flower color and the hypocotyls color (Groose and Palmer, 1991), but it was only reported in immature flowers. By searching the Public EST Project of Soybean in Soybase (<http://soybase.agron.iastate.edu>), there is no EST library constructed with normal purple hypocotyls.

Expression of the *DFR7* gene was found in many tissues including roots, hypocotyls, mature flowers, young seedlings, and shoots. It may be produced in the immature flower too, but the amount was too low to be detectable. This could be the reason why there are some anthocyanins synthesized in T322w, even though no *DFR* transcripts were observed.

## Acknowledgements

The authors would like to thank Dr. R.W. Groose, Department of Plant, Soil and Insect Sciences, University of Wyoming, Laramie, WY, for providing flower pictures, Dr. R.C. Shoemaker, USDA ARS and the Department of Agronomy, Iowa State University, Ames, IA, for providing the EST clone Gm-c1086-2103 [gi:17022852], and Dr. M.P. Scott, USDA ARS and the Department of Agronomy, Iowa State University, Ames, IA, for the guidance on the HPLC analysis.

## References

- Akada S, Dube SK (1995) Organization of soybean chalcone synthase gene clusters and characterization of a new member of the family. *Plant Mol Biol* 29: 189-199
- Akada S, Kung SD, Dube SK (1990a) Nucleotide sequence of one member of soybean chalcone synthase multi-gene family. *Nucleic Acids Res* 18: 3398
- Akada S, Kung SD, Dube SK (1990b) The nucleotide sequence of gene 3 of the soybean chalcone synthase multigene family. *Nucleic Acids Res* 18: 5899
- Akada S, Kung SD, Dube SK (1991) The nucleotide sequence of gene 1 of the soybean chalcone synthase multigene family. *Plant Mol Biol* 16: 751-752
- Akada S, Kung SD, Dube SK (1993a) Nucleotide sequence and putative regulatory elements of gene 2 of the soybean (*Glycine max*) chalcone synthase multigene family. *Plant Physiol* 102: 317-319
- Akada S, Kung SD, Dube SK (1993b) Nucleotide sequence and putative regulatory elements of a nodule-development-specific member of the soybean (*Glycine max*) chalcone synthase multigene family, *Gmchs 7*. *Plant Physiol* 102: 321-323
- Akada S, Kung SD, Dube SK (1993c) Nucleotide sequence of a soybean chalcone synthase gene with a possible role in ultraviolet-B sensitivity, *Gmchs6*. *Plant Physiol* 102: 699-701
- Brett D, Pospisil H, Valcarcel J, Reich J, Bork P (2002) Alternative splicing and genome complexity. *Nature Genet* 30: 29-30
- Bright SW, Lea PJ, Miflin BJ (1979) The regulation of methionine biosynthesis and metabolism in plants and bacteria. *Ciba Found Symp* 72: 101-117

- Burbulls IE, Lacobucci M, Shirley W (1996) A null mutation in the first enzyme of flavonoid biosynthesis does not affect male fertility in *Arabidopsis*. *Plant Cell* 8: 1013-1025
- Fasoula DA, Stephens PA, Nickell CD, Vodkin LO (1995) Cosegregation of purple-throat flower color with dihydroflavonol reductase polymorphism in soybean. *Crop Sci* 35: 1028-1031
- Groose RW, Palmer RG (1991) Gene action governing anthocyanin pigmentation in soybean. *J Hered* 82: 498-501
- Groose RW, Schulte SM, Palmer RG (1990) Germinal reversion of an unstable mutation for anthocyanin pigmentation in soybean. *Theor Appl Genet* 79: 161-167
- Groose RW, Weigelt HD, Palmer RG (1988) Somatic analysis of an unstable mutation for anthocyanin pigmentation in soybean. *J Hered* 79: 263-267
- Harborne JB (1958) Spectral methods of characterising anthocyanins. *Biochem J* 70: 22-28
- Hartwig EE, Hinson K (1962) Inheritance of flower color in soybeans. *Crop Sci* 2: 152-153.
- Holton TA, Cornish EC (1995) Genetics and biochemistry of anthocyanin biosynthesis. *Plant Cell* 7: 1071-1083
- Kato KK, Palmer RG (2003) Molecular mapping of male-sterile, female-sterile mutant gene (*st8*) in soybean. *J Hered* 94: 425-428
- Kato KK, Palmer RG (2004) Molecular mapping of four ovule lethal mutants in soybean. *Theor Appl Genet* 108: 577-585
- Keim P, Olson T, Shoemaker RC (1988) A rapid protocol for isolating soybean DNA. *Soybean Genet Newsl* 15: 150-152

- Lei HY, Zhou B, Zhang Y, Hong GF (2002) Structural analysis of a gene cluster encoding DFR-like proteins from rice chromosome 4. *Acta Biochim Biophys Sin* 34: 685-689.
- Maes T, De Keukeleire P, Gerats T (1999) Plant tagnology. *Trends Plant Sci* 4: 90-96
- McKeown M (1992) Alternative mRNA cplicing. *Annu Rev Cell Biol* 8: 133-155
- Nozzolillo C (1973) A survey of anthocyanin pigments in seedling legumes. *Can J Bot* 51: 911- 915
- Palmer RG, Groose RW (1993) A new allele at the *w4* locus derived from the *w4-m* mutable allele in soybean. *J Hered* 84: 297-300
- Palmer RG, Hedges BR, Benavente RS, Groose RW (1989) *w4*-mutable line in soybean. *Dev Genet* 10: 542-551
- Palmer, RG, Groose, RW, Weigelt, HD, Miller, JE (1990) Registration of a genetic stock (*w4-mw4-m*) for unstable anthocyanin pigmentation in soybean. *Crop Sci* 30: 1376-1377
- Ramachandran S, Sundaresan V (2001) Transposons as tools for functional genomics. *Plant Physiol Biochem* 39: 243-252
- Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular cloning: a laboratory manual*. 2nd edn. Cold Spring Harbor Laboratory Press. Plainview, NY
- Strack D (1997) Phenolic Metabolism. In: Dey PM & Harborne JB (eds), *Plant Biochemistry*. Academic Press, San Diego. pp: 387-416
- Thomas PS (1980) Hybridization of denaturated RNA and small DNA fragments transferred to nitrocellulose. *Proc Natl Acad Sci USA* 77:5201–5205
- Todd JJ, Vodkin LO (1996) Duplications that suppress and deletions that restore expression from a CHS multigene family. *Plant Cell* 8: 687-699

- Trick HN, Dinkins RD, Santarem ER, Di R, Samoylov V, Meurer C, Walker D, Parrott WA, Finer JJ, Collins GB (1997) Recent advances in soybean transformation. *Plant Tissue Cult Biotechnol* 3: 9-26
- Tuteja JH, Clough SJ, Chan WC, Vodkin LO (2004) Tissue-specific gene silencing mediated by a naturally occurring chalcone synthase gene cluster in *Glycine max*. *Plant Cell* 16: 819-835
- Walbot V (2000) Saturation mutagenesis using maize transposons. *Curr Opin Plant Biol* 3: 103-107
- Xu M, Palmer RG (2005) Genetic analysis and molecular mapping of a pale flower allele at the *W4* locus in soybean. *Genome* 48: 334 –340
- Xu YL, Li L, Wu K, Peeters AJ, Gage DA, Zeevaart JA (1995) The *GA5* locus of *Arabidopsis thaliana* encodes a multifunctional gibberellin 20-oxidase: molecular cloning and functional expression. *Proc Natl Acad Sci USA* 92: 6640-6644

Table 1. Sequences homologous to dihydroflavonol-4-reductase (DFR) gene in NCBI and GenBank database

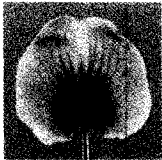
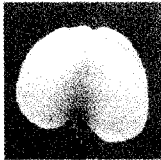
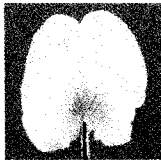
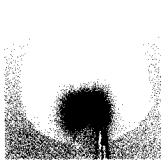

Name	Access number	GI number	Cultivar	Tissue	Homology to DFR1 gene
DFR1	AW185819	6455136	Williams	Immature seed coats	98%
	AW705650	7589885	Williams		96%
	AW705450	7589681	Williams		96%
	AW597470	7284983	Williams		99%
	AF167556	5852932	Chin-Ren-Woo-Dou		100%
	BM091941	17020907	Williams 82		96%
	BM091866	17020832	Williams 82		100%
DFR2	DQ026299	66356299	Harosoy	Immature flowers of field-grown plants	84%
	BE058910	8403276	Williams 82		81%
DFR3	BI893861	16106121	Williams 82	Etiolated hypocotyls	79%
DFR4	AI416884	4260388	Williams 82	na	79%
DFR5	CA802608	26059694	Williams	Immature pods (~2cm long) of greenhouse-grown plants	85%
DFR6	AW164620	6341734	T157	Seed coats of greenhouse-grown plants	94%

**Note:** na, not available; ns, no similarity.

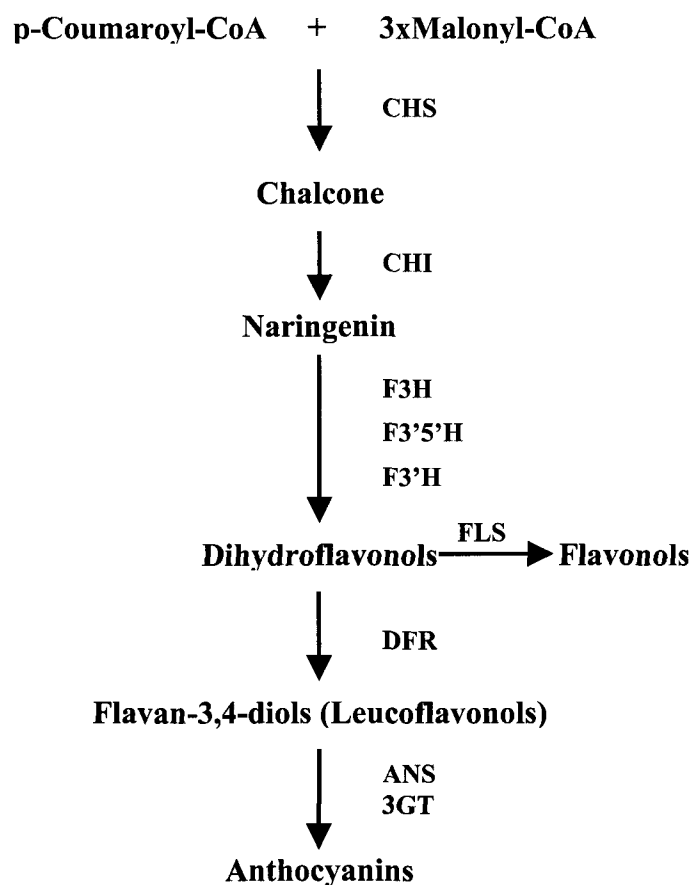


Table 1. Continued

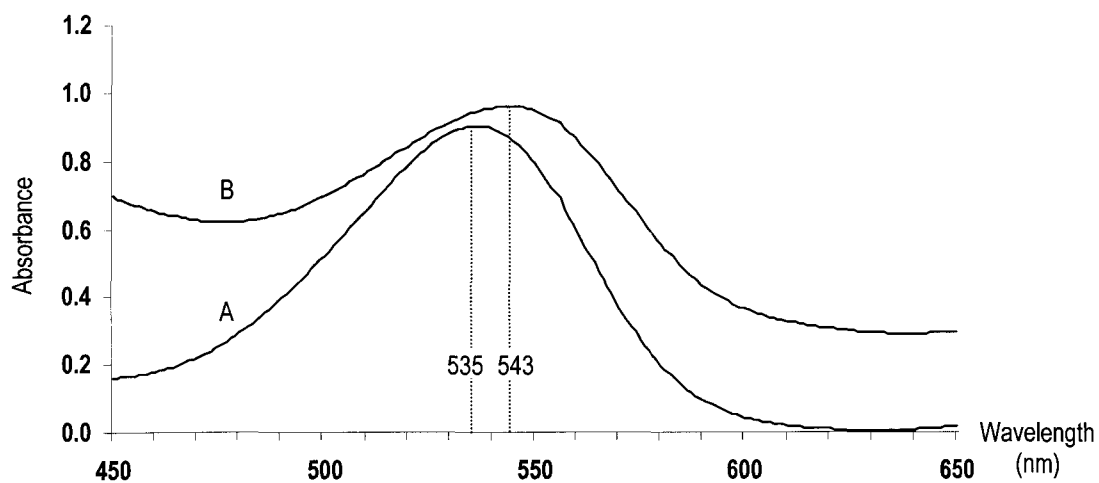
Name	Access number	GI number	Cultivar	Tissue	Homology to DFR1 gene
DFR7	AW156480	6227881	Williams 82	Mature flowers, field-grown plants	91%
	AW423450	6951382	Williams 82		94%
	BM885832	19269576	Williams	Germinating shoot, 24-hour germination	
	BQ081250	19937341	Williams 82	Etiolated hypocotyls	87%
	BQ080127	19935097	Williams 82		89%
	BQ079966	19934936	Williams 82		86%
	BG509046	13479703	Williams 82	Seedlings induced for symptoms of SDS disease	90%
	BG508964	13479621	Williams 82		90%
	BM526528	18731565	<i>Glycine soja</i>	Whole seedling, 2-week old, etiolated	85%
	BU083498	22524687	Clark	whole seedlings, 3-week old	86%
	BI699965	15662594	Bragg	Roots of 7-day old 'Bragg' seedlings	92%
	BQ473555	21285684	Minsoy x Noir1 RI progeny	Roots of bulked Minsoy x Noir RI progeny	86%
	AW233911	6566230	Supernod	Roots	86%
	AW234865	6567254	Supernod	Roots of 'Supernod' plants	83%
	AW310155	6725756	Supernod		90%
	AW309485.	6725086	Supernod		86%
	AW203651	6502278	Supernod		86%
	AW831783	7925757	Supernod		84%
	BM520810	18691962	<i>Glycine soja</i>	Whole seedling, 2-week old, etiolated	ns
	BF596687	11688949	Williams	Degenerating cotyledons	ns
	AI939277	5678147	Williams 82	Immature flowers of field-grown plants	ns

Soybean lines	Harosoy	Harosoy <i>w4</i>	T369	T321	T322
Genotype	<i>W4W4</i>	<i>w4w4</i>	<i>w4-pw4-p</i>	<i>w4-dpw4-dp</i>	<i>w4-mw4-m</i>
Flower color	Purple	Near white	Pale	Dilute purple	Variegated
Standard petals					

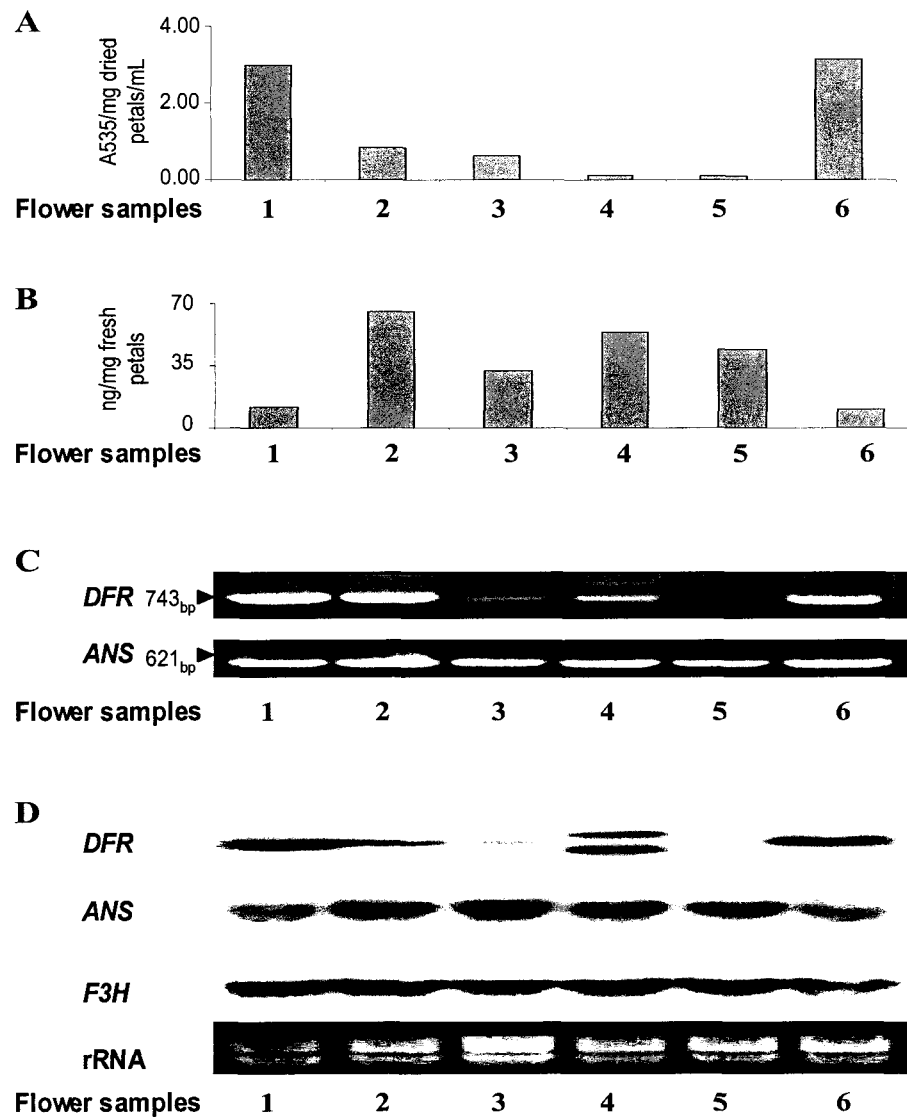
**Fig. 1** Soybean lines used in HPLC analysis and RNA analysis experiments.



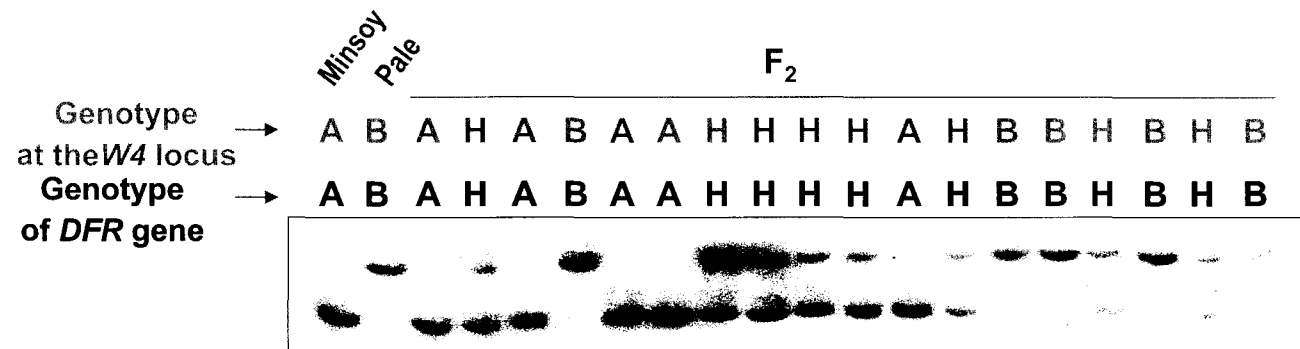
**Fig. 2** Anthocyanin biosynthesis pathway. CHS, Chalcone Synthase; CHI, Chalcone Isomerase; F3H, Flavanone 3-Hydroxylase; F3'5'H, Flavonoid 3'5'-Hydroxylase; F3'H, Flavonoid 3-Hydroxylase; DFR, Dihydroflavonol Reductase; ANS, Anthocyanin Synthase; 3GT, 3-Glucose Transferase; FLS, Flavonol synthase.



**Fig. 3** Absorption spectra of anthocyanins extracted from immature flowers of cv. Harosoy. **A.** Absorption spectrum of anthocyanin extracts in methanol-HCl, the absorption peak is at 535nm; **B.** Absorption spectrum of anthocyanin aglycones hydrolyzed by boiling 30 minutes, the absorption peak shifted 8 nm to 543nm.



**Fig. 4** Results from spectrophotometry analysis, HPLC, RT-PCR, and Northern blot analysis. Immature flowers of six samples from five soybean lines were analyzed. Flower sample 1, Harosoy (purple); 2, T369 (pale); 3, T321 (dilute purple); 4, Harosoy w4 (near white); 5, T322w (white section of variegated flower); and 6, T322p (purple section of variegated flower). **A.** Contents of total anthocyanins in six samples. **B.** Contents of myricetin in six samples. **C.** RT-PCR products of *DFR* and *ANS* genes from six samples. **D.** The transcript level of three genes in six samples determined by Northern blot analysis. They are *DFR*, *ANS*, and *F3H*.



**Fig. 5** RFLP mapping of DFR gene. Mapping population was F<sub>2</sub> generation from Minsoy (*W4W4*) X T369 (*w4-pw4-p*). DFR probe was generated by RT-PCR. Results showed that the DFR RFLPs co-segregated with the *W4* locus. The genotype at the *W4* locus of each F<sub>2</sub> plant was determined by progeny test in the F<sub>3</sub> generation (A: Homozygous Minsoy genotype; H: Heterozygous genotype; B: Homozygous T369 genotype).

## CHAPTER 4 GENETIC ANALYSIS OF FOUR NEW MUTANTS AT THE UNSTABLE *K2 MDH1-N Y20* CHROMOSOMAL REGION IN SOYBEAN [*GLYCINE MAX* (L.) MERR.]

A paper to be submitted to *Journal of Heredity*

Min Xu and Reid G. Palmer

### Abstract

In soybean, a chromosomal region defined by three closely linked genes, *k2*, *Mdh1-n*, and *y20*, is very unstable. A total of 31 mutants have been reported from this region. A mutation with tan-saddle seed was found from bulk-harvested seed of cultivar Kenwood. Genetics analysis established that the mutation is a recessive *k2* allele. SSR analysis showed the *k2* mutation is not a contamination from other existing *k2* mutants. Kenwood-*k2*, along with T239 (*k2*) and T261 (*k2 Mdh1-n*), were used to test for genetic instability at the *k2 Mdh1-n y20* chromosomal region by crossing with cv. Williams. Three new mutants were identified at the *k2 Mdh1-n y20* chromosomal region in crosses with T239. No mutants were found in crosses with T261 and Kenwood-*k2*. The new mutants provide additional genetic materials to study chromosomal aberrations in this unstable chromosomal region in soybean.

## Introduction

Several mutable or unstable loci have been described in soybean [*Glycine max* (L.) Merr.]. They include *Y18-m*, *cyt-Y3* (T284M), *w4-m*, *wp-m*, and *r-m* (for review, see Palmer et al. 2004).

A mutable *w4-m* allele has been reported at the *w4* locus, which is involved in the production of anthocyanin pigmentation (Groose et al. 1988). The *w4*-mutable line was proposed to contain an autonomous element at or near the *W4* locus (Palmer et al. 1989). The *w4*-mutable allele reverts at a high frequency from the recessive form (white) to the dominant form (purple) either somatically or germinally (Groose et al. 1990). Germinal reversion, in most cases, resulted in plants that produced all wild-type purple flower and hypocotyls. Many new mutations were identified among self-pollinated progenies of these wild-type germinal revertants (Palmer et al. 1989). Occasionally, germinal revertants may produce intermediate flower-color plants with green hypocotyls (Groose et al. 1990). One dilute-purple and one pale flower mutation were found. The dilute-purple mutation was an allele at the *W4* locus, and was assigned gene symbol *w4-dp*, and Genetic Type Collection number T321 (Palmer and Groose 1993). The pale mutation also was an allele at the *W4* locus, and was assigned gene symbol *w4-p*, and Genetic Type Collection number T369 (Xu and Palmer 2005). The *W4* locus has been mapped to soybean molecular linkage group D2 at 2.3 cM from the SSR markers Satt386 and close to the nearby telomere (Xu and Palmer 2005).

The *k2* locus conditions a tan saddle on the seed coat. Plants homozygous for a null allele, *Mdh1-n*, do not show malate dehydrogenase 1 (MDH1) activity on starch gel electrophoresis (Hedges and Palmer 1992). The *y20* locus conditions yellowish-green leaves.



The *k2 Mdh1-n y20* chromosome region is very unstable (Chen and Palmer 1998). So far 31 mutants have been reported from this region (Xu and Palmer 2005). A total of 25 *y20* mutants have been described, 18 are associated with *k2*, but all 25 are associated with *Mdh1-n*. No *k2k2 Mdh1Mdh1 y20y20* plants have been identified. Chen and Palmer (1998) have proposed that the instability at the *k2 Mdh1-n y20* chromosomal region was due to transposon activity that could generate chromosomal rearrangements such as deletions.

Imsande et al. (2001), using Southern blotting analysis, reported that the *Mdh1-n* mutants examined (T253, T317, T323, and T324) were the result of deletions. The null phenotype correlated with the deletion of specific genomic restriction fragments that encoded the *Mdh1* gene.

Five mapping populations were used by Xu and Palmer (unpublished results) to determine the order and position of *k2 Mdh1-n y20*. They also identified SSR markers in T261 that corresponded to deleted chromosome segments.

Chen and Palmer (1998) reported that tan-saddle mutant T239 was very effective in generating *Mdh1-n y20* mutants in certain cross-combinations. This “instability” experiment gave about 2.6% *Mdh1-n* mutants (10 out of 383 F<sub>2</sub> families) from crosses of T239 X parents of *w4-m* and *Y18-m*. The control population of 833 F<sub>2</sub> families from cross of T239 X cv. Harosoy gave no mutants. Tan-saddle mutant Clark-k2 (L67-3483) gave no mutants in 273 F<sub>2</sub> families. Genetic Type T261 (*k2 Mdh1-n*) gave about 0.3% *y20* mutants (1 mutant out of 323 F<sub>2</sub> families).

Tan-saddle mutant seed were identified in a bulk harvest of cv. Harosoy, precluding the identity of the original plant(s). The self-pollinated progeny of some tan-saddle seed were true-breeding and were green plants (T239), and some were true-breeding and were

yellowish-green foliage. The latter plants also were identified as malate dehydrogenase null (*Mdh1-n*) and were added to the soybean Genetic Type collection as T253 (*k2 Mdh1-n y20*). It was suspected that a transposon was responsible for the mutant phenotypes (Palmer 1984; Palmer et al., 1989). Genetic type T261 (*k2 Mdh1-n*) was found as a spontaneous mutation in cv Mandarin (Ottawa). The Clark-*k2* (L67-3843) tan saddle was found after seed X-irradiation of cv. Clark. Tan-saddle mutant seed were found in a bulk harvest of cv. Kenwood in 1993.

Our objectives were: 1) determine the inheritance and allelism of this tan-saddle mutant in Kenwood; and 2) to test for “instability” of Kenwood-*k2*, T239, and T261 in crosses with cv. Harosoy and Williams.

## **Materials and Methods**

The soybean lines, genotypes, and phenotypes used in the inheritance, allelism, and instability studies are given in Table 1.

### **Origin and inheritance of Kenwood-*k2***

Tan-saddle seed were found in a bulk harvest of cv. Kenwood (Cianzio et al. 1990) in 1993 at the Bruner Farm near Ames, Iowa. The plant with the most intense tan-saddle was single-plant threshed in 1994. Planting, selecting for intense tan saddle, and harvesting self-pollinated seed were continued for the next three years.

The Kenwood-*k2* plant with the most intense tan saddle was crossed as male parent with cv. Harosoy (Weiss and Stevenson 1955). The F<sub>1</sub> seed were advanced to the F<sub>2</sub> at the

University of Puerto Rico/ Iowa State University soybean nursery near Isabela. F<sub>2</sub> and F<sub>3</sub> seed were planted at the Bruner Farm near Ames, Iowa for seed-coat color evaluation.

### **Allelism and genetic instability studies with Kenwood-*k2***

For the allelism test, the Kenwood-*k2* plant with the most intense tan saddle was crossed as male parent to tan-saddle Clark-*k2* (L67-3483), which arose from X-irradiation of cv. Clark (Johnson 1958).

To test for instability at the *k2 Mdh1-n y20* chromosomal region, cv. Harosoy and Williams (Bernard and Lindahl 1972) were crossed reciprocally with Kenwood-*k2*. The F<sub>1</sub> seed from both tests were advanced to the F<sub>2</sub> at Isabela, Puerto Rico. F<sub>2</sub> and F<sub>3</sub> seed were planted at the Bruner Farm for evaluation of the presence /absence of tan-saddle seed and for seedling/adult-plant traits in the genetic instability study.

### **Test of origin of Kenwood-*k2* with SSR analysis**

SSR analysis was conducted as previously described (Xu and Palmer 2005). Six soybean lines including Kenwood, Kenwood-*k2*, T239, and Clark-*k2* were evaluated with 100 SSR markers, five from each molecular linkage group, respectively.

### **Test of instability of T239 and T261**

To test instability at the *k2 Mdh1-n y20* chromosomal region, Genetic Types T239 and T261 were crossed reciprocally with cv. Williams. The F<sub>1</sub> seed were advanced to the F<sub>2</sub> at Isabela, Puerto Rico. F<sub>2</sub> and F<sub>3</sub> seed were planted at the Bruner Farm for evaluation of the presence/absence of tan-saddle seed and for seedling/adult plant traits.

## Genetic evaluation of mutants

Three F<sub>2</sub> entries from crosses involving T239 and Williams segregated about 3 green: 1 yellow viable plant. Green and yellow F<sub>2</sub> plants were threshed individually and evaluated as plant-progeny rows.

Self-pollinated seed from the original yellow plants within each entry were evaluated for malate dehydrogenase according to the procedure of Cardy and Beversdorf (1984).

For inheritance studies, yellow plants from each of the three entries were crossed reciprocally with cv. Harosoy. Seed generation advance and data evaluation were similar to the instability study with T239 and T261.

For allelism tests, yellow plants from each of the three entries were crossed with T325 (*Mdh1-n y20*). A small piece of cotyledon was taken from each F<sub>1</sub> and F<sub>2</sub> seed. The samples were evaluated for malate dehydrogenase. The F<sub>1</sub> and F<sub>2</sub> plants were classified for plant color at the Bruner Farm.

## Results

### Inheritance of Kenwood-*k2*

Three hybrid seed were obtained from the cross of Harosoy X Kenwood-*k2*. The three F<sub>2</sub> progenies each segregated about 3 yellow seed coat: 1 tan-saddle seed coat (Table 2), which suggested that the mutation with tan-saddle seed coat found in cv. Kenwood was conditioned by a single gene and that it was recessive to wild-type yellow seed coat.

### **Allelism of Kenwood-*k2***

Tan-saddle Clark-*k2* was crossed with Kenwood-*k2*. The F<sub>1</sub> seed were advanced to the F<sub>2</sub> and all F<sub>2</sub> plants were tan-saddle seed. Twenty F<sub>2</sub> plants were evaluated as F<sub>3</sub> families the following summer at the Bruner Farm. Ten F<sub>3</sub> plants from each of the 20 progeny rows were evaluated for seed-coat color pattern. All plants were tan-saddle seed. This result established that tan-saddle-seed-coat mutations of Kenwood-*k2* and Clark-*k2* were allelic. They both were conditioned by the *k2* allele.

### **Instability studies with Kenwood-*k2***

Harosoy and Williams were crossed reciprocally with Kenwood-*k2* to test for instability at the *k2 Mdh1-n y20* chromosomal region. A total of 1705 F<sub>2</sub> progeny rows were evaluated six times from emergence to maturity (Table 3). The only variant observed was one foliage chimeric plant from the Kenwood-*k2* X Williams combination. Progeny of this plant were wild-type, no variation was observed.

### **SSR analysis of Kenwood-*k2***

One hundred SSR markers representing the 20 chromosomes of soybean genome were used to determine if Kenwood-*k2* was really a mutation from cv. Kenwood. No polymorphism between Kenwood-*k2* and Kenwood were detected by these 100 markers, but about 25% polymorphisms were found between Kenwood-*k2* and T239 or Clark-*k2*. This indicated that tan-saddle mutation of Kenwood-*k2* was not a contamination from other lines with the *k2* allele.

### **Origin of yellow plants from instability studies with T239 and T261**

Williams was crossed reciprocally with T239 and T261 to produce 216 F<sub>2</sub> families (Table 4). Three of 119 F<sub>2</sub> progenies from the cross with Williams and T239 were observed to segregate about 3 green: 1 yellow plants (Table 5), which is about 2.6% frequency. Green and yellow plants were threshed individually and progeny tested. The F<sub>2</sub> green plants gave about 1 all green: 2 segregating (green and yellow) progenies. Within segregating progenies the ratio was about 1: 2: 1 (Table 5). The yellow F<sub>2</sub> plants were true breeding.

### **Inheritance of the yellow plants**

Yellow plants were crossed reciprocally with cv. Harosoy. The F<sub>1</sub> plant color from all crosses was green which suggested that the yellow phenotype was the result of a nuclear mutation. The F<sub>2</sub> data and F<sub>2</sub>:<sub>3</sub> family data were in agreement with a single-gene recessive inheritance pattern (Table 6). The yellow F<sub>2</sub> plants were true breeding.

### **Allelism tests with T325**

The yellow plants were crossed with T325, a *Mdh1-n y20* mutant identified in a gene-tagging study with *w4-m*. The parents, F<sub>1</sub>, and a random sample of F<sub>2</sub> seed were tested for malate dehydrogenase, and were evaluated for plant color. The parents, F<sub>1</sub>, and F<sub>2</sub> plants were all yellow foliage (*y20*) and the seeds were malate dehydrogenase null (*Mdh1-n*) (Table 7). The three independently derived mutants are alleles of *Mdh1-n y20*.

## Discussion

In soybean, the recessive *k2* allele conditions tan-saddle seed coat color, while the dominant *K2* allele conditions yellow seed coat. A new mutant with tan-saddle seed was found in a bulk harvest of cv. Kenwood. According to the results from inheritance and allelism studies, the mutation was conditioned by a single locus, which is an allele at the *K2* locus. SSR analysis confirmed that the mutation was not a contamination from Clark-*k2* or T239 (*k2*).

Besides the *K2* locus, three loci including *I*, *R*, and *T* have major effects on the seed coat color (for review, see Palmer et al. 2004). The *I* locus is an inhibitor of the anthocyanin synthesis in seed coat. Soybean lines would produce the yellow seed coat with dominant *I* allele, and produce black seed coat with homozygous recessive *i* allele. The other two alleles at this locus, *i-i* and *i-k*, condition yellow seed with black hilum or black saddle seed, respectively. The intensity of the black color is modified by the *R* and *T* loci.

The *I* locus consists of two inverted repeat (IR) complexes. Each of them was a cluster of three chalcone synthase (CHS) genes (*CHS1*, *CHS3*, and *CHS4*) arranged as IRs in a 10 kb region (Todd and Vodkin 1996; Tuteja et al. 2004). The transcripts of these inverted CHS genes could form double-strand RNA (ds RNA), which could lead to degradation of homologous RNAs, and silence all the homologous CHS genes posttranscriptionally. Deletions in the promoter region of *CHS4* and *ICHS1* would release the gene silencing, increase the total transcript level of CHS genes, especially *CHS7/8*, and lead to black-pigmented seed coat (Tuteja et al. 2004).

The *k2* mutation and the *i-k* mutation lead to similar phenotypes. Both of them recover the lack of anthocyanins caused by low transcript level of CHS genes in a restricted

region of the seed coat (Wang et al. 1994; Tuteja et al. 2004), and result in colored saddle seed coat. How does the *k2* mutation recover the transcript level of CHS genes?

RFLP analysis showed that the soybean genome might go through more than one round of duplications (Shoemaker et al. 1996). So, the first hypothesis is that the *K2* locus is a duplicated region of the *I* locus and could inhibit the transcription of CHS genes. However, the CHS genes inhibited by the *K2* locus should be less homologous to the ones silenced by the *I* locus, so the *I* locus only has weak effects on them; otherwise, the *k2* mutation would not complement the inhibition by the *I* locus. In soybean, eight CHS genes from *CHS1* to *CHS8* had been cloned and sequenced (Akada et al., 1990a, 1990b, 1991, 1993a, 1993b, 1993c; Akada and Dube, 1995). The homology among them is very high. Tuteja et al. (2004) classified them into two subgroups. The first group included *CHS1* to *CHS6*, the homology between each is above 90%. The second group included *CHS7* and *CHS8*, the homology between them is 97%. And the homology between two subfamilies is more than 80%. However, there are still more CHS genes not yet analyzed, some of which may have less homology to the identified CHS genes. For example, by searching NCBI database, three EST clones from immature flowers were transcribed from a CHS gene that is different from all the eight genes above. The homology between this CHS gene and the other genes was just around 73%. This CHS gene also was found expressed in the root.

The second hypothesis is that the *K2* locus regulates the transcription of the CHS genes at the *I* locus. Mutation at the *K2* locus would reduce or turn off the transcription of the CHS genes at the *I* locus, release the PTGS triggered by the *I* locus, and recover the anthocyanin synthesis in seed coats.



In soybean, the *k2* locus was tightly linked to the *Mdh1-n* and *y20* loci and resides in a chromosomal segment that is a 'hotspot' for mutation. The mutations of *Mdh1-n* and *y20* are suspected to correspond to a chromosomal segment deletion. The instability was suspected to be due to an inactive transposon residing in the region or nearby (Chen and Palmer 1998). It was activated after transposase was provided from an active transposon *in trans*.

As we know, the insertion and excision of transposable elements would generate rearrangements of the sequences flanking their insertion sites including deletions, inversions, and duplications. Sometimes the deletion size could be very large. For example, the Tam3 transposable element has induced a deletion with size more than 20 kb at the *niv* locus in *Antirrhinum majus* (for review, see Martin and Lister 1989).

Three more mutants with malate dehydrogenase 1 null (*Mdh1-n*) and yellow foliage (*y20*) were found in the instability experiments by crossing T239 with cv. Williams. But no mutant was found by crossing T261 and Kenwood-*k2* with cv. Williams. The mechanism of the *k2* mutation may be different from the *Mdh1-ny20* mutation. No *k2* mutants have been found in gene tagging studies (Palmer et al. 1989). The *k2* mutation could result from some special chromosomal structures such as inverted repeats.

The new mutants would provide additional genetic materials for studying the chromosome aberration in this chromosomal region.

## Acknowledgement

We thank Dr. K. S. Lewis, USDA ARS, for providing the Kenwood-*k2* seeds.

## References

- Akada S and Dube SK, 1995. Organization of soybean chalcone synthase gene clusters and characterization of a new member of the family. *Plant Mol Biol* 29:189-199.
- Akada S, Kung SD, and Dube SK, 1990a. Nucleotide sequence of one member of soybean chalcone synthase multi-gene family. *Nucleic Acids Res* 18:3398.
- Akada S, Kung SD, and Dube SK, 1990b. The nucleotide sequence of gene 3 of the soybean chalcone synthase multigene family. *Nucleic Acids Res* 18:5899.
- Akada S, Kung SD, and Dube SK, 1991. The nucleotide sequence of gene 1 of the soybean chalcone synthase multigene family. *Plant Mol Biol* 16:751-752.
- Akada S, Kung SD, and Dube SK, 1993a. Nucleotide sequence and putative regulatory elements of gene 2 of the soybean (*Glycine max*) chalcone synthase multigene family. *Plant Physiol* 102:317-319.
- Akada S, Kung SD, and Dube SK, 1993b. Nucleotide sequence and putative regulatory elements of a nodule-development-specific member of the soybean (*Glycine max*) chalcone synthase multigene family, *Gmchs 7*. *Plant Physiol* 102:321-323.
- Akada S, Kung SD, and Dube SK, 1993c. Nucleotide sequence of a soybean chalcone synthase gene with a possible role in ultraviolet-B sensitivity, *Gmchs6*. *Plant Physiol* 102:699-701.
- Bernard RL and Lindahl DA, 1972. Registration of Williams soybean. *Crop Sci* 12:716.
- Cardy BJ and Beversdorf WD, 1984. Identification of soybean cultivars using isozyme electrophoresis. *Seed Sci Technol* 12: 943-954.
- Palmer RG and Chen XF, 1998. Instability at the *k2 Mdh1-n y20* chromosomal region in soybean. *Mol Gen Genet* 260:309-318.

- Cianzio SR, Schultz SP, Voss BK, and Fehr WR, 1990. Registration of 'Kenwood' soybean. *Crop Sci* 30:1162.
- Groose RW, Weigelt HD, and Palmer RG, 1988. Somatic analysis of an unstable mutation for anthocyanin pigmentation in soybean. *J Hered* 79:263-267.
- Groose RW, Schulte SM, and Palmer RG, 1990. Germinal reversion of an unstable mutation for anthocyanin pigmentation in soybean. *Theor Appl Genet* 79:161-167.
- Hedges BR and Palmer RG, 1992. Inheritance of malate dehydrogenase nulls in soybean. *Biochem Genet* 30:491-502.
- Imsande J, Pittig J, Palmer RG, Wimmer C, and Gietl C, 2001. Independent spontaneous mitochondrial malate dehydrogenase null mutants in soybean are the result of deletions. *J Hered* 92:333-338.
- Johnson HW, 1958. Registration of soybean varieties, VI. *Agron J* 50:690-691.
- Martin C and Lister C, 1989. Genome juggling by transposons: Tam3-induced rearrangements in *Antirrhinum majus*. *Dev Genet* 10:438-451.
- Palmer RG 1984. Pleiotropy or close linkage of two mutants in soybeans. *J Hered* 75:457-462.
- Palmer RG and Groose RW, 1993. A new allele at the *w4* locus derived from the *w4-m* mutable allele in soybean. *J Hered* 84:297-300.
- Palmer RG, Pfeiffer TW, Buss GR, and Kilen TC, 2004. Qualitative genetics. In: *Soybeans: Improvement, Production, and Uses*, 3rd ed (Boerma HR and Specht JE, eds). Madison, Wisconsin: ASA, CSSA, and SSSA; 137-214.
- Palmer RG, Hedges BR, Benavente RS, and Groose RW, 1989. *w4*-mutable line in soybean. *Dev Genet* 10:542-551.

- Shoemaker RC, Polzin K, Labate J, Specht J, Brummer EC, Olson T, Young N, Concibido V, Wilcox J, Tamulonis JP, Kochert G, and Boerma HR, 1996. Genome duplication in soybean (*Glycine* subgenus *soja*). *Genetics* 144:329-339.
- Todd JJ and Vodkin LO, 1996. Duplications that suppress and deletions that restore expression from a chalcone synthase multigene family. *Plant Cell* 8:687-699.
- Tuteja JH, Clough SJ, Chan WC, and Vodkin LO, 2004. Tissue-specific gene silencing mediated by a naturally occurring chalcone synthase gene cluster in *Glycine max*. *Plant Cell* 16:819-835.
- Wang C, Todd JJ, and Vodkin LO, 1994. Chalcone synthase mRNA and activity are reduced in yellow soybean seed coats with dominant *I* alleles. *Plant Physiol* 105:739-748.
- Weiss MG and Stevenson TM, 1955. Registration of soybean varieties V. *Agron J* 47:541-543.
- Xu M and Palmer RG, 2005. Genetic analysis and molecular mapping of a pale flower allele at the *W4* locus in soybean. *Genome* 48:334 -340.

Table1. Soybean lines and descriptions used in the instability studies

<b>Parents</b>	<b>Genotype</b>	<b>Phenotype</b>
Harosoy	<i>K2K2 Mdh1 Mdh1</i>	Wild type
Williams	<i>K2K2 Mdh1 Mdh1</i>	Wild type
Kenwood- <i>k2</i>	<i>k2k2</i>	Tan-saddle seed coat
Clark- <i>k2</i>	<i>k2k2</i>	Tan-saddle seed coat
T239	<i>k2k2</i>	Tan-saddle seed coat
T261	<i>k2k2 Mdh1-n Mdh1-n</i>	Tan-saddle seed coat and malate dehydrogenase 1 null

Table 2. Inheritance of tan-saddle seed coat from the cross of  
Harosoy X Kenwood-*k*2; F<sub>2</sub> data

Entry	Seed-coat patterns		$\chi^2_{(3:1)}$	P
	Yellow	Tan saddle		
A02-353-1	282	84	0.82	0.37
A02-353-2	319	100	0.29	0.59
A02-353-3	213	67	0.17	0.68
Total	814	251	1.16	0.28

Table 3. Test for instability of Kenwood-*k2* with  
cultivars Harosoy and Williams

<b>Entry</b>	<b>No. of F<sub>2</sub> families</b>
Harosoy X Kenwood- <i>k2</i>	322
Kenwood- <i>k2</i> X Harosoy	118
Williams X Kenwood- <i>k2</i>	601
Kenwood- <i>k2</i> X Williams	664
<b>Total</b>	<b>1705</b>

Table 4. Test for instability of T239 and T261 with cultivar Williams

<b>Cross</b>	<b>No. F<sub>2</sub> families</b>	<b>No. F<sub>2</sub> families seg. green and yellow plants*</b>
Williams X T239	15	1
T239 X Williams	104	2
Williams X T261	6	0
T261 X Williams	91	0

\*A00-R-78, A00-R-147, and A00-M-41 were the F<sub>2</sub> families segregating green and yellow plants.



Table 5. Inheritance of the three yellow mutants found in the test for instability of T239 with cultivar Williams

Family number*	No. F <sub>2</sub> plants				No. F <sub>2:3</sub> families				No. plants within segregating F <sub>2:3</sub> families			
	Green	Yellow	$\chi^2_{(3:1)}$	P	Green	Seg.	$\chi^2_{(1:2)}$	P	Green	Yellow	$\chi^2_{(3:1)}$	P
A00-R-78	127	43	0.01	0.93	39	85	0.20	0.66	2797	913	0.30	0.58
A00-R-147	51	18	0.04	0.83	18	30	0.37	0.54	1418	465	0.09	0.76
A00-M-41	72	26	0.12	0.73	23	42	0.12	0.73	1703	574	0.05	0.82

\*A00-R-78 and A00-R-147 were identified in different F<sub>2</sub> families from the cross of T239 X Williams.

A00-M-41 was identified in an F<sub>2</sub> family from the cross of Williams X T239.

Table 6. Inheritance of the three yellow mutants found in the test for instability of T239 and T261 crossed with Williams. Mutant plants were crossed with Harosoy

Parent 1*	Parent 2*	No. F <sub>2</sub> plants				No. F <sub>2:3</sub> families				No. plants within segregating F <sub>2:3</sub> families			
		Green	Yellow	$\chi^2_{(3:1)}$	P	Green	Seg.	$\chi^2_{(1:2)}$	P	Green	Yellow	$\chi^2_{(3:1)}$	P
A00-R-78	Harosoy	609	199	0.06	0.81	41	86	0.06	0.80	3093	1001	0.66	0.42
A00-R-147	Harosoy	735	238	0.15	0.70	44	95	0.18	0.67	5097	1686	0.07	0.78
A00-M-41	Harosoy	485	157	0.10	0.75	37	69	0.12	0.73	3052	1029	0.10	0.75

\*Data from reciprocal crosses bulked, homogeneity tests indicated that all samples were homogeneous.

Table 7. Allelism tests of the three yellow, malate dehydrogenase F<sub>2</sub> plants crossed with T325  
(*Mdh1-n y20*)

<b>Female</b>	<b>Male</b>	<b>No. F<sub>1</sub> plants</b>		<b>No. F<sub>2</sub> plants</b>	
		<b>Foliage color</b>	<b>MDH pattern</b>	<b>Foliage color</b>	<b>MDH pattern*</b>
A00-R-78	T325	10, yellow	10, null	369, yellow	30, null
A00-R-147	T325	12, yellow	12, null	381, yellow	36, null
A00-M-41	T325	14, yellow	14, null	411, yellow	42, null

\*Three F<sub>2</sub> seed descended from each of the F<sub>1</sub> plants were evaluated for malate dehydrogenase.

**CHAPTER 5 MOLECULAR MAPPING OF *K2 MDH1-N Y20*, AN  
UNSTABLE CHROMOSOMAL REGION IN SOYBEAN  
[*GLYCINE MAX* (L.) MERR.]**

**A paper submitted to Theoretical and Applied Genetics**

Min Xu and Reid G. Palmer

**Abstract**

In the soybean genome, a chromosomal region covering three tightly linked genes, *k2*, *Mdh1-n*, and *y20*, was found very unstable. It was suspected that the instability of the *k2 Mdh1-n y20* chromosomal region was caused by a non-autonomous transposable element residing adjacent to or in this region. In this study, we located and mapped this region with simple sequence repeat (SSR) markers on the soybean integrated map using five mapping populations. The *k2 Mdh1-n y20* chromosomal region was located on molecular linkage group (MLG) H. The integrated map from five mapping populations consisted of 13 loci in the order Satt541, Satt469, Sat\_122, Satt279, Satt253, Satt314, *Mdh1-n*, *y20*, *k2*, Satt302, Satt142, Satt181, and Satt434. The *k2 Mdh1-n y20* chromosomal region was very close to Satt314, Satt253, and Satt279. The genetic distance between the *Mdh1-n* gene and Satt314 was less than 1 cM. The results of the mapping study were consistent with the results from previous studies that the *Mdh1-n* mutation in T261 (*k2 Mdh1-n*) and the *Mdh1-n y20* mutation in T317 (*Mdh1-n y20*) were caused by deletions. In addition, another putative deletion was found in the genome of T261 which covered three SSR markers (Satt314, Satt253, and Satt279).

## Introduction

Three recessively inherited loci, *k2*, *Mdh1-n*, and *y20*, were tightly linked in the soybean genome. The *k2* locus conditions tan-saddle seed coat, the *Mdh1-n* locus conditions mitochondrial malate dehydrogenase 1 (MDH1) null, and the *y20* locus conditions chlorophyll-deficient foliage. The recombination rate between *k2* and *Mdh1-n* was detected as  $1 \pm 1.36\%$  in the cross of T261 (*k2 Mdh1-n*) X wild-type cultivars (*K2 Mdh1*) (Chen and Palmer 1996). The recombination rate between *k2* and *Mdh1-n y20* was zero in the coupling phase [T253 (*k2 Mdh1-n y20*) X Clark-*w1* (L69-4776) (*K2 Mdh1 Y20*)] (Palmer 1984), and  $3.04 \pm 0.48\%$  in the repulsion phase [T317 (*Mdh1-n y20*) X T239 (*k2*) or Clark-*k2* (*k2*)] (Chen and Palmer 1998a). But no recombination was found between *Mdh1-n* and *y20* in any of the three cross-combinations. The co-segregation between the *Mdh1-n* and *y20* loci is possibly caused by a chromosome deletion. Southern blot results using watermelon *Mdh* cDNA as a probe showed that all the *Mdh1-n y20* mutants tested were missing a 5.5 kb *EcoRI* band, which corresponded to soybean *Mdh1* gene [gi:5929963] (Pittig et al. 1994; Imsande et al. 2001).

The *k2 Mdh1-n y20* chromosomal region is very unstable (Chen and Palmer 1998b). So far, 31 mutants have been reported from this region (Table 1) (Palmer et al. 2004). Among these mutants, four chlorophyll deficient (CD) lines (T323, T324, T325, and T346) with *Mdh1-n y20* alleles were found in the descendents from germinal revertants of the *w4*-mutable line, T322, which was proposed to contain an active transposable element at the *W4* locus (Palmer et al. 1989; Chen et al. 1999). Two mutants with *Mdh1-n y20* alleles (T317 and

T361) were found from tissue culture-derived plants (Amberger et al. 1992; Palmer et al. 2000). Twelve mutants with *k2 Mdh1-n y20* alleles, T334, T335, T336, T337, T338, T339, T340, T341, T342, T343, T344, and T345, were identified in the F<sub>2</sub> descendants derived from crosses of T239 (*k2*) or T261 (*k2 Mdh1-n*) with the wild-type parental strains of the *w4-m* and *Y18-m* mutable lines that were suspected to contain an active transposon in their genomes (Chen and Palmer 1998b). The hypothesis for the instability of the *k2 Mdh1-n y20* chromosomal region is that a non-autonomous transposable element could reside adjacent to or in this region (Chen and Palmer 1998b). The non-autonomous transposable element could excise from its chromosomal position and cause mutations such as chromosome deletions, when it is activated by tissue culture or by transposases provided *in trans* through transposon tagging experiments or crossing experiments.

In the present study, the objective was to position the *k2 Mdh1-n y20* region in the soybean integrated genetic linkage map (Song et al. 2004) and to determine the order of the *k2*, *Mdh1-n*, and *y20* genes with simple sequence repeat (SSR) markers.

## Materials and Methods

### Plant material and DNA extraction

Five F<sub>2</sub> mapping populations named POP-1, -2, -3, -4, and -5, were used to map the *k2 Mdh1-n y20* chromosomal region. They were constructed from five different crosses, each derived from a single F<sub>1</sub> seed. The details are described in Table 2. F<sub>2</sub> seeds in POP-1 and -2 were placed on germination paper at 32°C in a dark growth chamber for 3 days. Two punches of samples were taken from cotyledons of 3-day-old seedlings by a 100 µl micropipette.

Samples were analyzed using starch gel electrophoresis to determine malate dehydrogenase isozyme banding patterns (Cardy and Beversdorf 1984). The seedlings were placed into a 32°C growth chamber with light, and transplanted in the field at the Bruner Farm (Ames, IA) after 4 days. All plants were recorded for seed-coat color at maturity and threshed individually. F<sub>2</sub> plants in POP-3, -4, and -5 were grown at the Bruner Farm. Leaf color was scored after 3 weeks, and seed-coat color was scored at maturity. Plants were threshed individually. The genotypes of F<sub>2</sub> individuals in POP-1, -3, and -5 were evaluated for the target gene (or genes) by analyzing their F<sub>2:3</sub> families.

About 3 to 4 grams young leaves were sampled from each F<sub>2</sub> plant and their parental lines of the five mapping populations, and ground into powders after they were freeze-dried. Genomic DNA was extracted from the leaf powders using CTAB method (Keim et al. 1988), diluted into 10 ng/μl, and stored at 4 °C.

### **Bulked segregant analysis (BSA)**

POP-1[cultivar Minsoy (PI 27890)(wild-type) X T261 (*k2 Mdh1-n*)] was used in bulked segregant analysis (BSA) (Michelmore et al. 1991) to identify the markers linked to the *k2 Mdh1-n* region. Two bulks were made. Bulk 1 (B1) contained DNA aliquots from 10 F<sub>2</sub> individuals in POP-1 homozygous for wild-type phenotypes (yellow seed coat and MDH1 present). Bulk 2 (B2) contained DNA aliquots from 10 F<sub>2</sub> individuals in POP-1 homozygous for tan-saddle seed coat and MDH1 null. SSR analysis was conducted with B1, B2, and the two parental lines (Minsoy and T261) to screen for candidate markers linked to the target region.

**SSR analysis**

SSRs were amplified by polymerase chain reaction (PCR) in 30  $\mu$ l mixture containing: 50 ng genomic DNA, 1x PCR buffer, 1.75 mM  $MgCl_2$ , 150  $\mu$ M dNTP, 0.15  $\mu$ M of each primer (Song et al. 2004), and 3 units *Taq* DNA polymerase (Promega, Madison, WI). The PCR was performed in a MJR PTC-100 thermal cycler (MJ Research, Inc., Waltham, MA) for 32 cycles of 45 s at 94 °C, 45 s at 47 °C, and 45 s at 68 °C. Amplified products were resolved and evaluated by electrophoresis on 2% Agarose 3:1<sup>TM</sup> (AMRESCO, Solon, OH) gels in 1x TBE (Tris/borate /EDTA) buffer, or on 8% (w/v) denaturing gels [29:1 acrylamide-bisacrylamide, 5.6 M urea, and 30% (v/v) formamide) in 1x TAE (Tris/acetate/EDTA) buffer.

**Linkage analysis**

The molecular linkage maps for five mapping populations first were calculated individually with Mapmaker 2.0 (Lander et al. 1987). The thresholds for linking two markers together were LOD 4.0 and recombination value 0.4. The genetic distances were converted from recombination rates using the Kosambi map function (Kosambi 1944). Then, the five maps were integrated with Joinmap 3.0 (Van Olijen and Voorrips 2001).



## Results

### Molecular mapping of the *k2 Mdh1-n* chromosomal region with mapping population POP-1

The mapping population POP-1 contained 118 F<sub>2</sub> individuals derived from a single F<sub>1</sub> seed from the cross of Minsoy X T261 (*k2 Mdh1-n*) (Table 2). The F<sub>2:3</sub> segregation for both the *k2* and *Mdh1-n* genes fit a 1 wild-type homozygote: 2 heterozygotes: 1 recessive homozygote ratio with  $\chi^2 = 0.02$ ,  $P = 0.99$  and  $\chi^2 = 0.02$ ,  $P = 0.99$ , respectively (Table 3).

BSA analysis was used to identify the SSR markers linked with the *k2 Mdh1-n* chromosomal region. A total of 182 SSR markers from the 20 soybean molecular linkage groups (MLGs) (Song et al. 2004), were selected initially to screen the two bulks (B1 and B2) constructed from POP-1 and the two parental lines (Minsoy and T261). The average genetic distances between any two adjacent markers were about 25 cM. The results showed that Satt253 in MLG H was polymorphic between the two parental lines and the two bulks. Its primers could amplify a ~138 bp band from Minsoy and B1, but none from T261 and B2, which indicated that the SSR marker Satt253 could be tightly linked with the *k2 Mdh1-n* chromosomal region. To confirm this, Satt253 was used to screen the entire POP-1 population. Since Satt253 could only be detected from Minsoy, it was a dominant marker in the POP-1 mapping population. The genotypes of homozygotes of Minsoy allele and heterozygotes could not be distinguished in F<sub>2</sub> generation. In both cases, a ~138 bp band could be amplified by the primers of Satt253. If “A” was used for homozygotes of Minsoy allele of one locus, “H” for heterozygotes of this locus, and “B” for homozygotes of this locus, “D” was used to represent either A or H when they could not be distinguished. The F<sub>2</sub>

D: B segregation ratio of Satt253 in POP-1 fit a 3: 1 ratio ( $\chi^2 = 0.01$ ,  $P = 0.92$ ) (Table 3). The result from Mapmaker 2.0 showed that Satt253 was linked to the *k2* locus with a LOD of 23.83 and the *Mdh1-n* locus with a LOD of 26.02.

An additional 14 markers, Satt353, Satt568, Satt192, Satt442, Satt541, Satt469, Sat\_122, Satt052, Satt314, Satt279, Satt302, Satt142, Satt181, and Satt434, from MLG H were used to screen the two parental lines of POP-1. These fourteen SSR markers and Satt253 also were used to screen the parental lines of the other four populations described in the following sections.

As a result, seven markers, Sat\_122, Satt279, Satt314, Satt302, Satt142, Satt181, and Satt434, showed polymorphism between the two parental lines, Minsoy and T261 (*k2 Mdh1-n*). Satt279 and Satt314 were dominant markers. Like Satt253, they can only be amplified from Minsoy, but not from T261. The others were co-dominant markers. The dominant markers segregated in POP-1 as a 3 D: 1B ratio, and the co-dominant markers segregated in POP-1 as a 1A: 2H: 1B ratio (Table 3). Data were subjected to analysis with Mapmaker 2.0. The most plausible position of the *k2 Mdh1-n* chromosomal region was located between Satt302 and three co-segregating SSR markers, Satt253, Satt279, and Satt314 with genetic distances of 18.9 cM (distance between *k2* and Satt302) and 0.8 cM (distance between *Mdh1-n* and Satt253, Satt279, and Satt314), respectively (Table 3). The *k2* gene was closer to Satt302, and the *Mdh1-n* gene was closer to the three co-segregating SSR markers. The recombination rate between *k2* and *Mdh1-n* was detected as 0.9%.

### **Molecular mapping of the *k2 Mdh1-n* chromosomal region with mapping population POP-2**

The mapping population POP-2 consisted of 122  $F_2$  plants derived from a single  $F_1$  seed of cultivar Williams X T261 (*k2 Mdh1-n*) (Table 2). The  $F_2$  segregation for both the *k2* and *Mdh1-n* genes fit a 3 wild-type: 1 homozygous recessive ratio with  $\chi^2 = 0.01$ ,  $P = 0.92$  and  $\chi^2 = 0.10$ ,  $P = 0.75$ , respectively (Table 4).

Eight markers that could detect polymorphism between the two parental lines were used generate the map. Dominant markers Satt314, Satt279, and Satt253 segregated in the mapping population as a 3: 1 ratio (Table 4). Co-dominant markers segregated for homozygous for the Williams allele: heterozygous: homozygous for the T261 (*k2 Mdh1-n*) allele in the mapping population as a 1: 2: 1 ratio (Table 4). The results showed that the *Mdh1-n* gene co-segregated with Satt314, Satt279, and Satt253. The *k2* gene was located between the *Mdh1-n* gene and the Satt302 marker with 0.8 cM genetic distance to the *Mdh1-n* gene, and 23.9 cM genetic distance to the Satt302 marker (Table 4).

### **Molecular mapping of the *k2* gene with mapping population POP-3 and -4**

The mapping population POP-3 was formed by crossing Williams with T239 (*k2*) (Table 2). The population contained 84  $F_2$  individuals, descendents from a single  $F_1$  seed. The genotype of each  $F_2$  individual was determined by  $F_{2:3}$  family analyses. The  $F_{2:3}$  segregation ratio for the *k2* gene fit a 1 wild-type homozygote: 2 heterozygotes: 1 recessive homozygote ratio ( $\chi^2 = 0.19$ ,  $P = 0.91$ ) (Table 5).

The mapping population POP-4 consisted of 113 F<sub>2</sub> plants derived from a single F<sub>1</sub> seed of Minsoy X Clark-*k2* (*k2*) (Table 2). The F<sub>2</sub> segregation for the *k2* gene fit a 3 wild-type: 1 homozygous recessive ratio ( $\chi^2 = 0.24$ ,  $P = 0.62$ ) (Table 6).

SSR markers that were polymorphic between the parental lines of each population were co-dominant, and segregated in the mapping populations as the ratio of 1A [homozygous for alleles from wild-types (Williams or Minsoy)]: 2H (heterozygous): 1B [homozygous for alleles from mutants (T239 or Clark-*k2*)] (Tables 5 and 6). The *k2* gene was mapped between Satt279 and Satt302 with 3.0 cM (POP-3) or 2.6 cM (POP-4) genetic distance to Satt279, and 24.5 cM (POP-3) or 24.0 cM (POP-4) genetic distance to Satt302 (Table 5 and 6)

#### **Molecular mapping of the *k2* and *y20* genes with mapping population POP-5**

The mapping population POP-5 was constructed from the cross of T317 (*Mdh1-n y20*) x T261 (*k2 Mdh1-n*) (Table 2). The population consisted of 84 F<sub>2</sub> individuals derived from a single F<sub>1</sub> seed. The genotype of each F<sub>2</sub> individual was determined by F<sub>2:3</sub> family analyses. The F<sub>2:3</sub> segregation ratio for both the *k2* and *y20* genes fit a 1 wild-type homozygote: 2 heterozygote: 1 recessive homozygote ratio with  $\chi^2 = 0.86$ ,  $P = 0.65$  and  $\chi^2 = 1.21$ ,  $P = 0.55$ , respectively (Table 7).

Seven out of 15 SSR markers detected polymorphisms between T317 and T261. Satt314, Satt253, and Satt279 were dominant markers since they could be amplified only from T317. These three markers segregated in the mapping population as a 3: 1 ratio (Table 7). The remaining four markers, Sat\_122, Satt302, Satt142, and Satt181 were co-dominant. The segregation pattern was 1 homozygous for the T317 (*Mdh1-n y20*) allele: 2

heterozygous: 1 homozygous for the T261 (*k2 Mdh1-n*) allele in the mapping population (Table 7). The recombination rate between the *k2* gene and the *y20* gene was 0.6%. The *k2* and *y20* genes were mapped between Satt302 and three co-segregating markers, Satt314, Satt253, and Satt279. The *y20* gene was closer to the three co-segregating markers with 2.8 cM genetic distance between each other, and the *k2* gene was closer to Satt302 with 21.6 cM genetic distance to Satt302 between each other.

### **Construction of an integrated map for the five mapping populations**

The data from the five mapping population were loaded into Joinmap 3.0 (Van Olijen and Voorrips 2001) to make an integrated map for the *k2 Mdh1-n y20* chromosomal region. The program first created a map for each mapping population. The gene order of each map was the same as the corresponding maps created by Mapmaker 2.0 (data not shown). Then the maps were combined by calculating the mean recombination frequencies and combined LOD scores of each pair of loci from all five mapping populations as described by Stam (1993). The integrated map contained 13 loci in the order; Satt541, Satt469, Sat\_122, Satt279, Satt253, Satt314, *Mdh1-n*, *y20*, *k2*, Satt302, Satt142, Satt181, and Satt434 (Fig. 1A).

## **Discussion**

### **Position of the *k2 Mdh1-n y20* chromosomal region on the soybean genetic map**

Simple sequence repeats (SSRs) are PCR-based markers. They can be very easily detected through standard PCR reactions. In soybean, SSR markers are usually multiple alleles, single locus, and inherited as co-dominant alleles (Akkaya et al. 1992). The

integrated genetic linkage map (Cregan et al. 1999) includes 606 SSRs loci, 689 RFLPs (Restriction Fragment Length Polymorphism), 79 RAPDs (Random Amplified Polymorphic DNA), and 11 AFLPs (Amplified Fragment Length Polymorphism) distributed on 20 molecular linkage groups which are assumed to represent the 20 chromosomes of the soybean genome, and covers about 2400 cM. Recently, a new integrated genetic linkage map of the soybean was released (Song et al. 2004). It included a total of 1849 markers, in which 1015 markers were SSRs, and covered about 2500cM of the chromosome genome. Thus, SSR markers are a good choice for positioning new genes on the soybean linkage map. By using SSR markers, we have mapped several genes, such as female partial-sterile genes *Fsp1* (Kato and Palmer 2003a), *Fsp2*, *Fsp3*, *Fsp4*, and *Fsp5* (Kato and Palmer 2004a), male-sterile and female-sterile gene *st8* (Kato and Palmer 2003b), lethal yellow gene *y18* (Kato and Palmer 2004b), and the mutable gene *w4-m* (Xu and Palmer 2005).

The primary goals of this study were to locate the *k2 Mdh1-n y20* chromosomal region in the soybean integrated genetic linkage map (Song et al. 2004) and to determine the order of the *k2*, *Mdh1-n*, and *y20* loci by SSR markers. Since the *k2*, *Mdh1-n*, and *y20* loci are very closely linked and no recombination between the *Mdh1-n* and *y20* loci was ever detected in previous genetics studies (Palmer 1984; Chen and Palmer 1996; Chen and Palmer 1998), their order on the chromosome could not be determined using only one mapping population. Thus five mapping populations POP-1, -2, -3, -4, and -5 were constructed to map the *k2* and *Mdh1-n* loci, the *k2* locus, and the *k2* and *y20* loci, respectively. Then data from these five populations were combined to generate one integrated map for the three loci.

The individual maps of the five mapping populations were created by Mapmaker 2.0 (Lander et al. 1987). The results were very consistent. All the overlapping loci in the five

mapping populations were in the same order on all the maps. The *k2 Mdh1-n y20* chromosomal region was located on MLG H, and tightly linked with three SSR markers Satt279, Satt253, and Satt314. The *Mdh1-n* and *y20* loci were positioned between the *k2* locus and the SSR markers (Tables 3, 4, 5, 6, and 7). The maps were integrated by Joinmap 3.0 (Van Olijen and Voorrips 2001). The result was basically consistent with the maps from Mapmaker 2.0. Although there was no recombination ever found between the *Mdh1-n* locus and *y20* locus in experiments, the genetic distance between these two loci was estimated as 0.83 cM by Joinmap 3.0 according the results from five mapping populations. The order of the three loci on the chromosome was determined as *Mdh1-n*, *y20*, and *k2* (Fig. 1A).

#### **Double chromosomal deletions in T261 (*k2 Mdh1-n*)**

Previous genetics analyses (Palmer 1984; Chen and Palmer 1998a) and Southern analysis (Pittig et al. 1994; Imsande et al. 2001) indicated that in soybean, most of the *Mdh1-n* or *Mdh1-n y20* mutations, including T261 (*k2 Mdh1-n*) and T317 (*Mdh1-n y20*), were caused by a chromosome deletion(s). The results from the mapping studies are in agreement with these results. The mapping experiments showed that the *Mdh1-n* and *y20* loci were located between the *k2* locus and Satt279, Satt253, and Satt314 (Tables 3, 4, 5, 6, and 7; Fig. 1A). With the *Mdh1-n* (T261) or *Mdh1-n y20* (T317) mutations (Tables 3, 4, and 7), the genetic distance between the *k2* locus and Satt279 was shorter. It was 1.7 cM, 0.8 cM, or 1.9 cM in mapping populations POP-1, -2, or -5, while 3.0 cM or 2.6 cM in POP-3 or -4 (Tables 3, 4, 5, 6, and 7). As the genetic distance is usually positively related to the physical distance, the shortness of genetic distance between the *k2* locus and Satt279 in POP-1, -2, and -5 was

assumed to be the result of the physical deletions at the *Mdh1-n* (T261) or *Mdh1-n y20* (T317) loci.

Besides the putative *Mdh1-n* deletion, MLG H in T261 may have another deleted region that covers three SSR markers (Satt253, Satt279, and Satt314). Firstly, in T261, there were no SSR alleles for these three SSR markers. That is, the SSR alleles could not be amplified from T261. Secondly, these three SSR markers co-segregated in the mapping populations using T261 as a parent (POP-1, -2, and -5 in Tables 3, 4, and 7), but not in the other mapping populations (POP-3 and -4 in Tables 5 and 6). These two phenomena indicated that the chromosomal region which covered Satt253, Satt279, and Satt314 may have been deleted. And, this deleted region was separate from the *Mdh1-n* deletion. If they were on the same deleted chromosomal segment, there would be no recombination between the SSR markers and the *Mdh1-n* locus because genes on the same deleted segment would co-segregate. However, the mapping results showed that there were meiotic crossovers between the SSR markers and the *Mdh1-n* or *Mdh1-n y20* deletions (Tables 1 and 7). So, in T261, there were two closely linked chromosome deletions on MLG H. The size of the deleted region covering three SSRs was estimated as approximately 1Mb according to the positions of the three SSRs on the physical map by Southern Illinois University, Carbondale (<http://bioinformatics.siu.edu/cgi-bin/gbrowse/SoyV4R4>). The size of the *Mdh1-n* deletion in T261 or the *Mdh1-n y20* deletion in T317 could not be determined in this study.



### **Comparisons of the integrated molecular map of the *k2 Mdh1-n y20* chromosomal region and the existing maps of linkage groups with *Mdh* genes**

In this study, the final map of the *k2 Mdh1-n y20* chromosomal region was made with Joinmap 3.0, which integrated all the segregation data from the five mapping populations. The integrated map was similar to the MLG H maps of Cregan et al. (1999) and Song et al. (2004). Interestingly, at the chromosomal location where we placed the *k2*, *Mdh1-n*, and *y20* loci (MLG H), there was a *Mdh* gene already placed on the USDA/Iowa State University map (Cregan et al. 1999) (Fig. 1 B). Is the *Mdh* gene on the USDA/Iowa State University map the same as the *Mdh1-n* gene, or is it a different *Mdh* gene that is linked to the *Mdh1-n* gene? To answer these questions, we checked the MDH isozyme immigration patterns on a 12% starch gel of the two parental lines used in the mapping population for constructing the original USDA/Iowa State University map, PI 468.916 (*G. soja*) and A81-356022 (*G. max*) (Shoemaker and Olsen 1993) (Fig. 2 Lanes 5 and 6).

Both parents were MDH1 positive, having the 1<sup>st</sup> and 2<sup>nd</sup> bands. However, PI 468.916 (*G. soja*) was MDH-A pattern, having the 6<sup>th</sup> band counting from the cathode, while A81-356022 (*G. max*), was the MDH-B pattern, missing the 6<sup>th</sup> band. MDH-A / MDH-B phenotypes were controlled by a single locus, with the MDH-A variant dominant to the MDH-B variant (Palmer et al. 1992). Here we use *Mdh-A* to represent the allele that conditions the MDH-A pattern, and *Mdh-B* to represent the allele that conditions the MDH-B pattern. Given these phenotypes, the *Mdh* gene on the USDA/Iowa State University map could not be the *Mdh1/Mdh1-n* gene but must be the *Mdh-A/B* gene, and according to the location it was placed on the USDA/Iowa State University map, it should be closely linked to the *Mdh1-n* gene on MLG H.

However, the same MDH-A/B polymorphism was detected between Harosoy and Clark, two parental lines of the mapping population used by the University of Nebraska (Fig. 2 Lanes 3 and 4), but in that population the *Mdh-A/B* gene was mapped on MLG D2 (Cregan et al. 1999) (Fig. 1 C). Many experiments supported this result. For example, a bacterial pustule resistant gene, *Rxp*, which was identified on the same classical linkage group with the *Mdh-A/B* gene (Palmer et al. 1992), was mapped onto MLG D2 (Narvel et al. 2001) with the position close to the *Mdh-A/B* gene mapped by the University of Nebraska (Fig. 1 C and D).

The reason could be that the *Mdh-A* genes located on different MLGs had different origins. The *Mdh-A* allele positioned on MLG H was from *G. soja*, whereas, the *Mdh-A* allele placed on MLG D2 was from *G. max*. It is inferred that the *Mdh-A* genes in *G. soja* and *G. max* are located on different chromosomes. However, they could be orthologous or paralogous loci, since they condition the same phenotype. But the results of comparative mapping showed that there was no synteny or duplicate loci between MLG D2 and MLG H (Shoemaker et al. 1996; Soybase: <http://129.186.26.94/>). We think this is possibly due to the absence of RFLP markers around *Mdh-A* genes on MLG D2 and H (Cregan et al. 1999; Song et al. 2004). So, to verify if these two *Mdh-A* genes are orthologous or paralogous loci, either new RFLP markers need to be developed, or more molecular work on gene cloning and sequencing should be done.

## Acknowledgements

The authors would like to thank Dr. R.C. Shoemaker, USDA ARS and the Department of Agronomy, Iowa State University, Ames, IA, for providing the seeds of *G.*

*soja* (PI 468.916) and A81-356022. The authors would also like to thank Dr. D. Grant, USDA ARS and the Department of Agronomy, Iowa State University, Ames, IA, for guidance in constructing maps using Mapmaker2.0.

## References

- Akkaya MS, Bhagwat AA, Cregan PB (1992) Length polymorphism of simple sequence repeat DNA in soybean. *Genetics* 132:1131-1139
- Amberger LA, Shoemaker RC, Palmer RG (1992) Inheritance of two independent isozyme variants in soybean plants derived from tissue culture. *Theor Appl Genet* 84:600-607
- Burzlaff JD, Palmer RG (1999) Soybean linkage studies: *y18* and *y20*. *Soybean Genet Newslett* 26 [Online journal] URL <http://www.soygenetics.org/articles/sgn1999-011.html> (posted 23 April 1999)
- Cardy BJ, Beversdorf WD (1984) A procedure for the starch gel electrophoretic detection of isozymes in soybean [*Glycine max* (L.) Merr.]. *Dep Crop Sci Tech Bull* 119-8401. Univ of Guelph, ON
- Chen XF, Imsande J, Palmer RG (1999) Eight new mutants at the *k2 Mdh1-n y20* chromosomal region in soybean. *J Hered* 90:339-403
- Chen XF, Palmer RG (1996) Inheritance and linkage with the *k2* and *Mdh1-n* loci in soybean. *J Hered* 87:433-347
- Chen XF, Palmer RG (1998a) Recombination and linkage estimation between the *k2* and *Mdh1-n y20* loci in soybean. *J Hered* 89:488-494

- Chen XF, Palmer RG (1998b) Instability at the *k2 Mdh1-n y20* chromosomal region in soybean. *Mol Gen Genet* 260:309-318
- Cregan PB, Jarvik T, Bush A, Shoemaker RC, Lark KG, Kahler A, Kaya N, VanToai T, Lohnes DG, Chung J, Specht JE (1999) An integrated genetic linkage map of the soybean genome. *Crop Sci* 39:1464-1490
- Imsande J, Pittig J, Palmer RG, Wimmer C, Gietl C (2001) Independent spontaneous mitochondrial malate dehydrogenase null mutants in soybean are the result of deletions. *J Hered* 92:333-338
- Kato KK, Palmer RG (2003a) Genetic identification of a female partial-sterile mutant in soybean. *Genome* 46:128-134
- Kato KK, Palmer RG (2003b) Molecular mapping of male-sterile, female-sterile mutant gene (*st8*) in soybean. *J Hered* 94:425-428
- Kato KK, Palmer RG (2004a) Molecular mapping of four ovule lethal mutants in soybean. *Theor Appl Genet* 108:577-585
- Kato KK, Palmer RG (2004b) Duplicate chlorophyll-deficient loci in soybean. *Genome* 47:190-198
- Keim P, Olson TC, Shoemaker RC (1988) A rapid protocol for isolating soybean DNA. *Soybean Genet Newslett* 15:150-152
- Kosambi DD (1944) The estimation of map distance from recombination values. *Ann Eugen* 12:172-175
- Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1:174-181

- Michelmore RW, Paran I, Kesseli RV (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proc Natl Acad Sci USA* 88:9828-9832
- Narvel JM, Jakkula LR, Phillips DV, Wang T, Lee SH, Boerma HR (2001) Molecular mapping of *Rxp* conditioning reaction to bacterial pustule in soybean. *J Hered* 92:267-270
- Palmer RG (1984) Pleiotropy or close linkage of two mutants in soybeans. *J Hered* 75:445-447
- Palmer RG, Burzlaff JD, Shoemaker RC (2000) Genetic analyses of two independent chlorophyll-deficient mutants identified among the progeny of a single chimeric foliage soybean plant. *J Hered* 91:297-303
- Palmer RG, Hedges BR, Benavente RS, Goose RW (1989) The *w4* mutable line in soybean. *Dev Genet* 10:542-551
- Palmer RG, Lim SM, Hedges BR (1992) Testing for linkage between the *Rxp* locus and nine isozyme loci in soybean. *Crop Sci* 32:681-683
- Palmer RG, Pfeiffer TW, Buss GR, Kilen TC (2004) Qualitative genetics. In Boerma HR, Specht JE (eds) *Soybeans: Improvement, Production, and Uses*, 3rd edn. ASA, CSSA, SSSA, Madison, WI, pp 137-214
- Pittig J, Guggenberger C, Palmer RG, Imsande J, Gietl C (1994) Molecular analysis of a deletion-prone chromosome segment in soybean [abstract 67]. Fourth international congress of plant molecular biology. Amsterdam: International Society of Plant Molecular Biology

- Shoemaker RC, Olson TC (1993) Molecular linkage map of soybean (*Glycine max* L. Merr.). In O'Brien SJ (ed.) Genetic Maps: Locus maps of complex genomes. Cold Spring Harbor Laboratory Press, New York, pp 6.131-6.138
- Shoemaker RC, Polzin K, Labate J, Specht J, Brummer EC, Olson T, Young N, Concibido V, Wilcox J, Tamulonis JP, Kochert G, Boerma HR (1996) Genome duplication in soybean (*Glycine* subgenus *soja*). Genetics 144:329-339
- Song QJ, Marek LF, Shoemaker RC, Lark KG, Concibido VC, Delannay X, Specht JE, Cregan PB (2004) A new integrated genetic linkage map of the soybean. Theor Appl Genet 109:122-128
- Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer Package: JoinMap. Plant J 3:739-744
- Van Ooijen JW, Voorrips, RE (2001) JoinMap® 3.0, software for the calculation of genetic linkage maps. Plant Research International, Wageningen, The Netherlands
- Xu M, Palmer RG (2005) Genetic analysis and molecular mapping of a pale flower allele at the *W4* locus in soybean. Genome 48:334-340

Table 1. Summary of soybean lines mutated at the *k2 Mdh1-n y20* chromosomal region

Mutant genes	<i>k2</i>	<i>k2 Mdh1-n</i>	<i>Mdh1-n</i>	<i>Mdh1-n y20</i>	<i>k2 Mdh1-n y20</i>	
Mutant lines	T239 (L63-365)	T261/S56-26	PI 567.391	T234	T253	T342/M-11-7
	Clark- <i>k2</i> (L67-3483)		PI 567.630A	T317/LA45-1-5-1	T334/X-197	T343/M-15-23
			Mandell	T323/CD-1	T335/X-203	T344/M-19-3
				T324/CD-2	T336/X-217	T345/M-20-11
				T325/CD-3	T337/X-219	T347/X-193
				T346/CD-9	T338/X-241	T348/X-194
				T361	T339/X-451	T349/RP-95-649
					T340/M-7-2	T350
					T341/M-11-4	T351

Table 2. The F<sub>2</sub> populations for mapping the k2 Mdh1-n y20 chromosomal region

Population	Parental lines	Genotype	Description	Population size	Target gene(s)
POP-1*	Minsoy ( ♀ )	<i>K2K2 Mdh1Mdh1 Y20Y20</i>	wild-type cultivar	118	<i>k2, Mdh1-n</i>
	T261 ( ♂ )	<i>k2k2 Mdh1-nMdh1-n Y20Y20</i>	tan-saddle seed coat and MDH1 null		
POP-2	'Williams ( ♀ )	<i>K2K2 Mdh1Mdh1 Y20Y20</i>	wild-type cultivar	122	<i>k2, Mdh1-n</i>
	T261 ( ♂ )	<i>k2k2 Mdh1-nMdh1-n Y20Y20</i>	tan-saddle seed coat and MDH1 null		
POP-3*	Williams ( ♀ )	<i>K2K2 Mdh1Mdh1 Y20Y20</i>	wild-type cultivar	84	<i>k2</i>
	T239 ( ♂ )	<i>k2k2 Mdh1Mdh1 Y20Y20</i>	tan-saddle seed coat		
POP-4	Minsoy ( ♀ )	<i>K2K2 Mdh1Mdh1 Y20Y20</i>	wild-type cultivar	113	<i>k2</i>
	Clark- <i>k2</i> ( ♂ )	<i>k2k2 Mdh1Mdh1 Y20Y20</i>	tan-saddle seed coat		
POP-5*	T317 ( ♀ )	<i>K2K2 Mdh1-nMdh1-n y20y20</i>	yellow seed coat, MDH1 null, and yellow foliage	84	<i>k2, y20</i>
	T261 ( ♂ )	<i>k2k2 Mdh1-nMdh1-n Y20Y20</i>	tan-saddle seed coat, MDH1 null, and green foliage		

\*Populations whose F<sub>2</sub> individuals had genotype-evaluation in the F<sub>2:3</sub> generation.



Table 3. Segregation of *k2*, *Mdh1-n*, and linked SSR markers in POP-1 derived from Minsoy X T261 (*k2 Mdh1-n*)

Locus	Distance <sup>a</sup> (cM)	Segregation <sup>b</sup>					$\chi^2_{(1:2:1)}$	$\chi^2_{(3:1)}$	<i>P</i>
		A	H	D	B	Total			
Sat_122	7.1	31	57		30	118	0.15		0.93
Satt314	1.7			89	29	118		0.01	0.92
Satt279	1.7			89	29	118		0.01	0.92
Satt253	1.7			89	29	118		0.01	0.92
<i>Mdh1-n</i>	0.9	29	59		30	118	0.02		0.99
<i>k2</i>	0	30	59		29	118	0.02		0.99
Satt302	18.9	36	52		30	118	2.27		0.32
Satt142	24.1	38	52		28	118	3.36		0.19
Satt181	28.4	37	54		27	118	2.54		0.28
Satt434	45.6	37	58		23	118	3.36		0.19

<sup>a</sup> Distance from the *k2* locus

<sup>b</sup> **A**: homozygous for the allele from Minsoy at this locus; **B**: homozygous for the allele from T261 at this locus; **H**: heterozygous for the allele at this locus; **D**: A or H. The genotypes of the *Mdh1-n* and *k2* loci were determined in the F<sub>2:3</sub> generation. The genotypes of SSR markers were determined in the F<sub>2</sub> generation.

Table 4. Segregation of *k2*, *Mdh1-n*, and linked SSR markers in POP-2 derived from Williams x T261 (*k2 Mdh1-n*)

Locus	Distance <sup>a</sup> (cM)	Segregation <sup>b</sup>					$\chi^2_{(1:2:1)}$	$\chi^2_{(3:1)}$	<i>P</i>
		A	H	D	B	Total			
Satt541	7.0	33	58		30	121	0.36		0.84
Sat_122	4.9	33	55		32	120	0.85		0.65
Satt314	0.8			90	32	122		0.10	0.75
Satt279	0.8			90	31	121		0.02	0.87
Satt253	0.8			90	32	122		0.10	0.75
<i>Mdh1-n</i>	0.8			90	32	122		0.10	0.75
<b><i>k2</i></b>	0.0			91	31	122		0.01	0.92
Satt302	23.9	30	58		34	122	0.56		0.76
Satt142	32.6	33	59		29	121	0.34		0.84
Satt181	37.3	39	54		28	121	3.40		0.18

<sup>a</sup> Distance from the *k2* locus

<sup>b</sup> **A**: homozygous for the allele from Williams at this locus; **B**: homozygous for the allele from T261 at this locus; **H**: heterozygous for the allele at this locus; **D**: A or H. The genotype of each locus was determined in the F<sub>2</sub> generation.

Table 5. Segregation of *k2* and linked SSR markers in POP-3 derived from Williams X T239 (*k2*)

Locus	Distance <sup>a</sup> (cM)	Segregation <sup>b</sup>				$\chi^2_{(1:2:1)}$	<i>P</i>
		A	H	B	Total		
Satt541	14.0	19	43	22	84	0.26	0.88
Sat_122	11.0	19	42	23	84	0.38	0.83
Satt314	4.2	22	44	18	84	0.57	0.75
Satt253	3.6	22	42	20	84	0.10	0.95
Satt279	3.0	23	41	20	84	0.26	0.88
<b><i>k2</i></b>	0.0	20	44	20	84	0.19	0.91
Satt302	24.5	17	44	23	84	1.05	0.59
Satt142	36.1	18	42	24	84	0.86	0.65

<sup>a</sup> Distance from the *k2* locus

<sup>b</sup> **A**: homozygous for the allele from Williams at this locus; **B**: homozygous for the allele from T239 at this locus; **H**: heterozygous for the allele at this locus. The genotype of the *k2* locus was determined in the F<sub>2:3</sub> generation. The genotypes of SSR markers were determined in the F<sub>2</sub> generation.

Table 6. Segregation of *k2* and linked SSR markers in POP-4 derived from Minsoy XClark-*k2*

Locus	Distance <sup>a</sup> (cM)	Segregation <sup>b</sup>					$\chi^2_{(1:2:1)}$	$\chi^2_{(3:1)}$	<i>P</i>
		A	H	D	B	Total			
Satt541	9.4	26	59		28	113	0.29		0.86
Satt469	9.4	26	59		28	113	0.29		0.86
Sat_122	6.7	25	61		27	113	0.79		0.67
Satt253	3.5	27	60		26	113	0.45		0.80
Satt279	2.6	28	60		25	113	0.59		0.74
<b><i>k2</i></b>	0.0			87	26	113		0.24	0.62
Satt142	24.0	27	55		31	113	0.36		0.83
Satt181	30.4	24	59		30	113	0.86		0.65
Satt434	52.3	24	51		38	113	4.54		0.10

<sup>a</sup> Distance from the *k2* locus

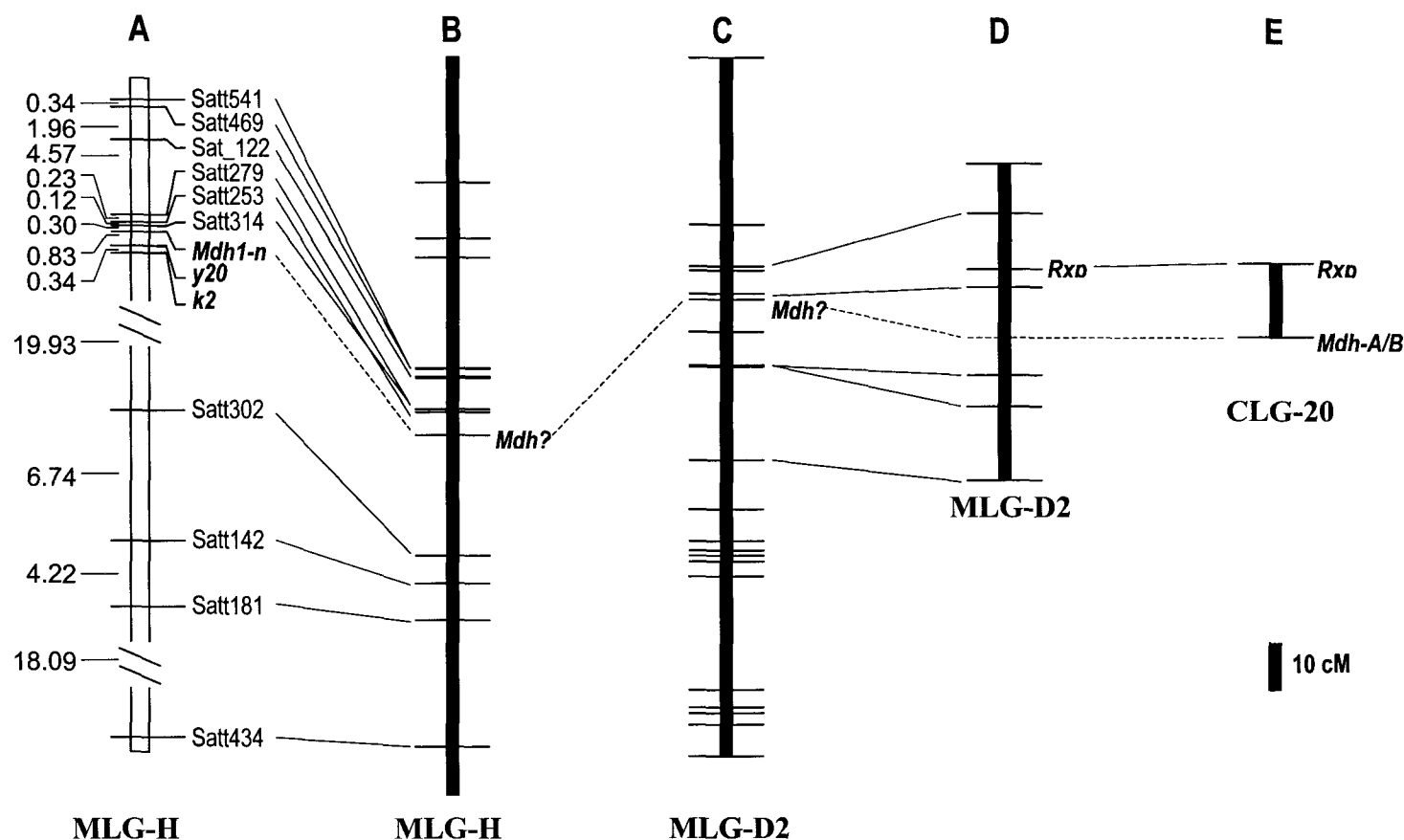
<sup>b</sup> **A**: homozygous for the allele from Minsoy at this locus; **B**: homozygous for the allele from Clark-*k2* at this locus; **H**: heterozygous for the allele at this locus; **D**: A or H. The genotype of each locus was determined in the F<sub>2</sub> generation.

Table 7. Segregation of *k2*, *y20*, and linked SSR markers in POP-5 derived from T317*(Mdh1-n y20)* x T261 (*k2 Mdh1-n*)

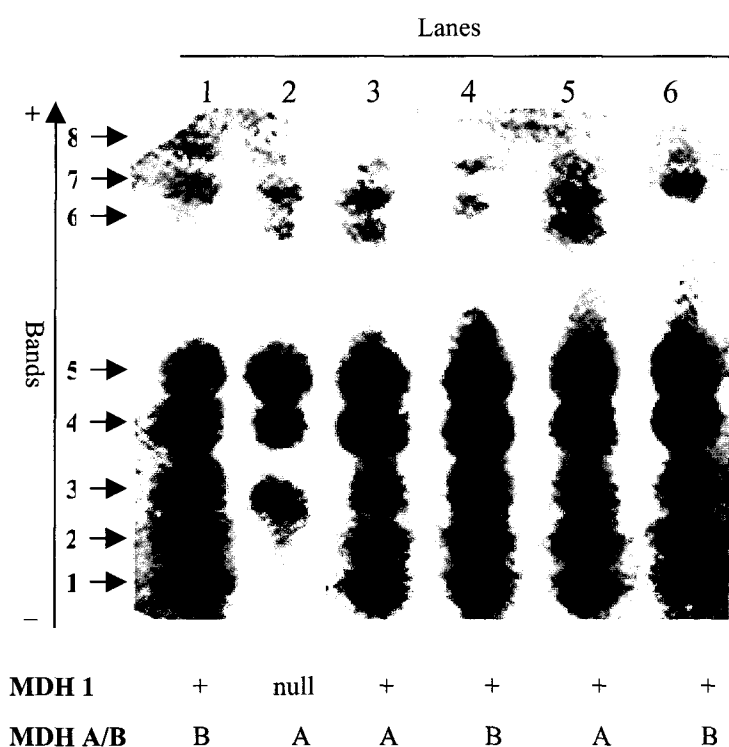
Locus	Distance <sup>a</sup> (cM)	Segregation <sup>b</sup>					$\chi^2_{(1:2:1)}$	$\chi^2_{(3:1)}$	<i>P</i>
		A	H	D	B	Total			
Sat_122	13.6	21	47		16	84	1.79		0.41
Satt253	1.3			67	17	84		1.02	0.31
Satt279	1.3			67	17	84		1.02	0.31
Satt314	1.3			67	17	84		1.02	0.31
<i>y20</i>	0.6	19	47		18	84	1.21		0.55
<i>k2</i>	0.0	20	46		18	84	0.86		0.65
Satt302	21.6	25	35		24	84	2.36		0.31
Satt142	28.4	25	40		19	84	1.05		0.59
Satt181	32.0	21	42		21	84	0.00		1.00

<sup>a</sup> Distance from the *k2* locus

<sup>b</sup> **A**: homozygous for the allele from T317 at this locus; **B**: homozygous for the allele from T261 at this locus; **H**: heterozygous for the allele at this locus; **D**: A or H. The genotypes of the *k2* and *y20* loci were determined in the F<sub>2:3</sub> generation. The genotypes of SSR markers were determined in the F<sub>2</sub> generation.



**Fig. 1** Comparison of five linkage maps. Distances are shown in centiMorgans (cM). The alleles between two maps were connected with straight lines, and the putative alleles between two maps were connected with dotted lines. A. The integrated molecular map for the *k2* *Mdh1-n y20* unstable chromosomal region constructed in this study. B. MLG H from the USDA/ Iowa State University map (Cregan et al. 1999). C. MLG D2 from the University of Nebraska map (Cregan et al. 1999). D. Molecular map of BP (bacterial pustule) resistant gene *Rxo* (Narvel et al. 2001). E. Classical linkage group 20 (Palmer et al. 1992).



**Fig. 2** The MDH patterns of six soybean lines on a 12% starch gel. Samples were run from the cathode to the anode. The mutant lines with MDH 1 null pattern were missing two MDH bands, band 1 and band 2. The soybean lines with MDH pattern A had band 6, while the soybean lines with MDH pattern B were missing band 6. **Lane 1.** Minsoy (*Mdh1Mdh1 Mdh-BMdh-B*); **Lane 2.** T261 (*Mdh1-nMdh1-n Mdh-AMdh-A*); **Lane 3.** Harosoy (*Mdh1Mdh1 Mdh-AMdh-A*); **Lane 4.** Clark(*Mdh1Mdh1 Mdh-BMdh-B*); **Lane 5.** PI 468.916 (*Glycine soja*) (*Mdh1Mdh1 Mdh-AMdh-A*); **Lane 6.** A81-356022 (*G. max*) (*Mdh1Mdh1 Mdh-BMdh-B*).

## CHAPTER 6 A LARGE DELETION IS INVOLVED IN THE *MDH1-N Y20* MUTATION IN SOYBEAN [*GLYCINE MAX* (L.) MERR.]

Min Xu and Reid G. Palmer

### Abstract

The *Mdh1-n* and *y20* loci along with the *k2* locus comprise an unstable chromosomal region in soybean. The *Mdh1-n* locus conditions mitochondrial malate dehydrogenase (MDH, EC 1.1.1.37) 1 null, and the *y20* locus conditions chlorophyll-deficient foliage. All the double mutations at the *Mdh1-n* and *y20* loci and some single mutations at the *Mdh1-n* locus were related to genomic deletions, and the cause of the deletions was suspected to be an inactive transposon residing nearby or within this region. The objectives in the present study were to identify the sizes and breakpoints of the deletions. To do that, an *Mdh1* contig with size of ~96 kb was constructed from three BAC clones 45N1, 146B8, and 203G2. Through Southern analysis, deletions in all the examined mutants (T261, T253, T317, T323, T325, and T361) were larger than ~70 kb. The breakpoints of the deletions could not be identified at this time. Three copies of truncated *Calypso5-1* like retroelements were found downstream of the *Mdh1* gene (AF180335), and clustered in a ~16 kb chromosomal region. To verify if these tandemly residing retroelements could play a role in the instability of the *Mdh1-n y20* region, it is necessary to identify the deletion breakpoints by screening a library with a large insertion size.



## Introduction

The *Mdh1-n* and *y20* loci along with the *k2* locus comprised an unstable chromosomal region in soybean [*Glycine max* (L.) Merr.]. The *k2* locus conditions tan-saddle seed coat, the *Mdh1-n* locus conditions mitochondrial malate dehydrogenase 1 (MDH1, EC 1.1.1.37) null, and the *y20* locus conditions chlorophyll-deficient foliage. The *k2 Mdh1-n y20* chromosomal region was mapped on molecular linkage group (MLG) H of the soybean integrated map (Song et al. 2004) with simple sequence repeat (SSR) markers and five mapping populations (Xu and Palmer data unpublished). It covered about 2 cM chromosomal region

No recombination was found between the *Mdh1-n* locus and *y20* locus in more than 30,000 F<sub>2</sub> plants from multiple crosses (Palmer 1984; Chen and Palmer 1996, 1998a). It was proposed that the co-segregation was due to a deletion. Southern blot analysis showed a 5.5 kb *EcoRI* band containing the 3'-end of soybean *Mdh1* gene [gi:5929963] existed in wild-type *Mdh1 Y20* soybean lines (Harosoy, Jilin 3, and T239) but was missing in all the *Mdh1-n y20* mutants tested (T253, T317, T323, and T324), using water melon mMDH cDNA as a probe (Imsande et al. 2001).

The *k2 Mdh1-n y20* chromosomal region is very unstable (Chen and Palmer 1998b). So far, 31 mutants have been reported at this region (for review, see Palmer et al. 2004). Of which, 25 mutants carried the *Mdh1-n y20* mutation, which were respectively generated by spontaneous, transposon tagging (Palmer et al. 1989; Chen et al. 1999), tissue culture (Amberger et al. 1992; Palmer et al. 2000), and out crossing in the instability experiments (Chen and Palmer 1998b). Four more mutants were found in this region (Xu and Palmer unpublished data). One *k2* mutation was spontaneously generated from cv. Kenwood. Three

*k2 Mdh1-n y20* mutants were isolated from the F<sub>2</sub> generation of crosses of T239 (*k2*) with cv. Williams (wild-type). It was suspected that a non-autonomous transposable element resided within or adjacent to the *k2 Mdh1-n y20* chromosomal region and made the region unstable (Chen and Palmer 1998b).

Recently, it was found that in *Arabidopsis*, excision of a hybrid *Dissociation* (*Ds*) element could induce a large genomic deletion flanking *Ds* (Page et al. 2004). It involved transposase mis-recognizing the *Ds* elements, using one end of newly transposed element and one end of the *Ds* transposon at the donor site as substrate, and excising out the region from these two ends.

To confirm the hypothesis that an inactive transposon residing within or nearby this region causes the instability of the *k2 Mdh1-n y20* chromosomal region, the identification of the break points of the deletions in the *Mdh1-n y20* mutants is very important. To do this, one needs to screen a genomic library with a large-size insertion. Bacterial artificial chromosome (BAC) vectors have been used to clone human DNA with inserts as large as 300 kb (Shizuya et al. 1992) and plant such as sorghum (Woo et al. 1994) with inserts averaging 157 kb. In soybean, many genes were cloned using BAC libraries (Lewers et al. 2002; Ashfield et al. 2003).

In the present study, the objectives were to identify the sizes and break points of the deletions in the *Mdh1-n y20* mutants.

## Materials and Methods

### Plant Materials

Soybean lines used are listed in Table 1. Cultivars Williams and Jilin 3 (PI 427.099) and Genetic Type T322 are wild-type cultivars for the *k2 Mdh1-n y20* region. T322 is a mutable line with purple and white variegated flowers and is suspected to contain an active transposable element (Groose et al. 1988). Genetic Types T323 and T325 with *Mdh1* null and yellow foliage (*y20*) mutants were isolated in a gene-tagging experiment using the *w4-m* mutable line T322 (Palmer et al. 1989). Genetic Types T317 and T361 also with *Mdh1* null and yellow foliage (*y20*) were found from tissue culture-derived plants of Jilin 3 (Amberger et al. 1992; Palmer et al. 2000). Genetic Types T253 (*k2 Mdh1-n y20*) and T261 (*k2 Mdh1-n*) are two spontaneous mutants found in the cultivars Harosoy and Mandarin (Ottawa), respectively.

Their Southern-blotting patterns with *EcoRI* digestion and watermelon *mMdh* cDNA as probe are shown on Fig. 1A.

### Identifying PCR based markers specifically corresponding to the *Mdh1* gene

Eight pairs of primers designed from the *Mdh1* gene sequence (AF180335) were used to screen wild-type soybean lines and the *Mdh1-n* or *Mdh1-n y20* mutants (Table 1) using polymerase chain reaction (PCR) method. They covered a 6466 bp chromosome region from 3008 bp upstream of the start codon of the *Mdh1* gene to the 3' end of the *Mdh1* gene. The details are listed in Table 2.

For confirmation, the putative marker(s) were used to screen 30 F<sub>2</sub> plants randomly selected from a mapping population made from Williams X T261 that was used to map the *k2* and *Mdh1-n* gene on the soybean integrated map (Song et al. 2004).

PCRs were performed in 50 µL of reaction mixture containing 1x PCR buffer, 2 mM MgCl<sub>2</sub>, 150 µM each dNTP, 0.25 µM each primer, 100 ng template DNA, and 3 units *Taq* DNA polymerase (Promega, Madison, WI). The profile of PCR consisted of initial denatured at 94 °C for 2 minutes, 32 cycles of denatured at 94 °C for 45 seconds, annealed at 58 °C for 45 seconds, and extended at 72 °C for 1 minute, and finished with extra extension at 72 °C for 10 minutes. The PCR products were evaluated by electrophoresis on 1% agarose gels in 1x TAE (Tris/acetate/EDTA) buffer.

#### **BAC library screening and contig construction**

A BAC library constructed from *Hind* III-digested cv. Williams 82 with 10-fold equivalents of soybean genome was kindly provided by Dr. Madan K. Bhattacharyya.

The library was primarily screened using watermelon mMDH cDNA as probe that was kindly provided by Dr. Christine Gietl. Positive clones were confirmed by PCR amplification with *Mdh1* specific primers and Southern blot analysis. Two ends of each confirmed positive clone were sequenced with primers U (21M13, 5'-TgT-AAA-ACg-ACG-gCC-AgT-3') and R (M13, 5'-CAg-gAA-ACA-gCT-ATg-ACC-3') respectively by using ABI 3730 DNA analyzer at the ISU DNA facility (Ames, IA). The insertion size of each clone was estimated from the sizes of the major fragments resulting from *Hind* III digestion and *Eco*RI digestion, respectively.

Each end of the positive clones was PCR-amplified using primers designed according to the sequencing results (Table 2). Each end amplified by PCR was used as probes to hybridize against each *Hind* III-digested clone to determine the correct alignment of positive clones by Southern analysis (Sambrook et al. 1989).

### **Deletion detection**

Deletion in mutants (Table 1) was detected with Southern blot analysis by comparison with the wild-type original lines (Table 1). Plant genomic DNA was extracted with CTAB method (Keim et al. 1987) and followed by purification with phenol-chloroform extraction. Ten micrograms DNA was digested with *Hind* III and separated on 0.8% agarose gel. The Southern blot analysis was conducted as previously described and detected with <sup>32</sup>P-labeled probes corresponding to the ends of positive BAC clones (Sambrook et al. 1989).

## **Results and Discussion**

### **Identifying PCR based markers specifically corresponding to the *Mdh1* gene**

A total of 8 pairs of primers (Table 2) were designed according to the *Mdh1* sequence (AF180335) (Imsande et al. 2001) for screening nine soybean lines that included three wild-types and six *Mdh1-n* or *Mdh1-n y20* mutants (Table 1). They covered a 6466bp chromosome region from the 3008<sup>th</sup> bp upstream of the start codon of the *Mdh1* gene to the 3' end of the *Mdh1* gene.

The primers of *Mdh1*-5' were designed for amplifying a 1426 bp segment that covered a chromosomal region from the start codon of the *Mdh1* gene to the beginning of the

second exon. The *Mdh1*-5' fragment could only be amplified from the wild-type soybean lines but not from the mutants with the *Mdh1-n* allele that was caused by deletion according to the Southern blot analysis (Fig. 1B).

In addition, *Mdh1*-5' was used to screen 30 F<sub>2</sub> plants from a mapping population made from Williams X T261. All the F<sub>2</sub> individuals with *Mdh1* allele (homozygous or heterozygous) could generate the 1426 bp band, while all the F<sub>2</sub> individuals with homozygous recessive *Mdh1-n* allele could not (data not shown).

All these results proved that marker *Mdh1*-5' was specific for the *Mdh1* gene. The other seven pairs of primers amplified same products from all the genetic materials, no matter whether they had wild-type *Mdh1* gene (*Mdh1*) or not (*Mdh1-n*) (the result from *Mdh1*-3' shown in Fig. 1C). This indicated that there were other *Mdh* genes highly homologous to the *Mdh1* gene in the soybean genome. Imsande et al. (2001) cloned a partial *Mdh* gene named as *Mdh2* (AF068686) that was 94% identical to the *Mdh1* gene in the coding and downstream regions.

### **Three BAC clones found containing the *Mdh1* gene**

A BAC library covered 10-fold of soybean genome size and constructed from cv. Williams 82 was probed with watermelon mMDH cDNA.

Sixteen positive clones after primary screening were examined using PCR method with the primers of *Mdh1*-5' to see if it included the *Mdh1* gene. Three of them generated a 1426 bp band, and they are 45N1, 146B8, and 203G2 (Fig. 2A). The PCR products were sequenced, and the results showed they are from the *Mdh1* gene. This is consistent with the results from Southern analysis. These 16 clones were digested with *EcoRI* and hybridized

with the probe used to screen the BAC library. The 5.5 kb band corresponding to the *Mdh1* gene was identified from clones 45N1, 146B8, and 203G2 (Fig. 2B). There were other two clones that had a band at the position of 8 kb that should correspond to other *Mdh* genes. No bands were identified from the remaining eleven clones (Fig. 2B). They could be false positive clones or picked incorrectly.

### **Constituting the *Mdh1* contig**

The ends of three positive clones (45N1, 146B8, and 203G2) were sequenced with ABI analyzer using primers U and R described in methods. New primers were designed to amplify each end of the clones based on the sequences. The primer information was listed in Table 3. Except for the reverse end of clone 45N1 (45N1R), all the others were amplified using the designed primers, which included the reverse ends of clones 146B8 and 203G2, named as 146B8R and 203G2R respectively, and the forward ends of clones 45N1, 146B8, and 203G2 named as 45N1F, 146B8F, and 203G2F, respectively.

To align these three clones, these five ends were used as probes to hybridize each clone with *Hind* III digestion. The results are shown in Fig. 3A. Probe 45N1R could only detect a 6 kb band from *Hind* III-digested clone 45N1, which indicated 45N1R should be one end of the contig. Probe 146B8R hybridized to two clones, 45N1 and 146 B8, and it could detect two bands from clone 146B8, and three bands from clone 45N1, which indicated this probe should be arranged outside of clone 203G2 but inside of clone 45N1. Probe 203G2F hybridized to all three clones, which means it was included in all three clones. Sequencing results showed that 146B8F and 203G2R were identical. Both of them hybridized to clones 146B8 and 203G2, but not to clone 45N1, which indicated they should be positioned at the

other end of the contig. Thus, the ends of three clones in the *Mdh1* contig should follow this order: 146B8F and 203G2R, 45N1F, 203G2F, 146B8R, and 45N1R (Fig. 3B).

According to the sizes of major bands from *Hind* III digested clones and *Eco*RI digested clones (data not shown), the insertion of clone 203G2 was about 80 kb, clone 146B8 was about 86 Kb, and clone 45N1 was about 72 kb. Compared with the band pattern of *Hind* III-digested clone 146B8, a 6 kb and a 3.8 kb fragments found in *Hind* III-digested clone 45N1 did not exist in the *Hind* III-digested clone 146B8 (data not shown). These two fragments should reside at the reverse end of clone 45N1 that is not overlapped by clone 146B8. Therefore, the size of this *Mdh1* contig comprised of clones 45N1, 146B8, and 203G2 should be around 96 Kb. The estimated physical position of each end of the clones in the contig is presented in Fig. 3B.

### **Retrotransposons residing downstream of the *Mdh1* gene**

The sequences of all six ends from three clones 45N1, 146B8, and 203G2 were blasted against NCBI non-redundant database using BLASTN program.

The results showed that the forward end of clone 203G2 (203G2F) shared 97% identity with the downstream sequence of the *Mdh1* sequence (AF180335). Through BLAST2 program, the forward end of clone 203G2 started from the 23629<sup>th</sup> nucleotide of the *Mdh1* sequence (AF180335) with an orientation toward the *Mdh1* gene whose coding region was from 7254 to 10704.

The reverse end of clone 146B8 (146B8R) was found homologous to the *pol* gene of *G. max* retrovirus-like element *Calypso5-1* from 5023 to 5661 (AF186186, Wright and Voytas 2002). The alignment of these two sequences was done with Pairwise global



alignment tool (EMBOSS) (Fig. 4). The homology was calculated as 77.8 %. The transcription direction of the *Calypso5-1* like element corresponding to 146B8R is opposite to the one of the *Mdh1* gene.

Southern analysis showed that there were three copies of *Calypso5-1* like elements residing in the *Mdh1* contig composed of clones 203G2, 146B8, and 45N1 (Fig. 3A). More accurately, they resided in a 16 kb chromosomal region from the 5' end of 203G2F to the 5' end of 45N1R. Two resided in the chromosomal region from the 5' end of 203G2F to the 5' end of 146B8 R, and one resided between the 5' ends of clones 146b8 and 45N1.

The *Calypso* elements in soybean are degenerated endogenous retroviruses (Wright and Voytas 2002). The coding regions of 24 elements cloned by Wright and Voytas (2002) were replete with stop codons, frameshifts, deletions, and insertions. The three copies of *Calypso5-1* like retroelements residing in a ~16 kb region of the *Mdh1* contig could be truncated fragments, since the length of *Calypso5-1* like elements is usually ranging from 12 to 14 kb (Wright and Voytas 2002). Similar structures were found in a BAC clone that corresponded to heterochromatin region in soybean chromosomes (Lin et al. 2005). Fragments homologous to the *pol* domain of *Calypso5-1* elements were dispersed within this BAC.

LTR-retrotransposon amplification plays a major evolutionary role in plant genome expansion in addition to polyploidization (Vicent et al. 1999; Meyers et al. 2001, Fu and Dooner 2002). To avoid “genomic obesity”, some mechanism would exist to remove retroelement DNA (Bennetzen 2002). In *Arabidopsis*, LTR-retrotransposons were mainly removed through illegitimate recombination (Devos et al. 2002), whereas, in rice, LTR-retrotransposons were removed through both unequal homologous recombination and

illegitimate recombination (Ma et al. 2004). It was found that in rice, insertions of most of LTR retrotransposons happened 8 million years ago. Now these elements have lost over two-thirds of their encoded sequences through deletions (Ma et al. 2004).

The truncated *Calypso 5-1* like retroelements residing downstream of the *Mdh1* gene in tandem could be another reason for the instability of the *Mdh1* gene in addition to the proposed inactive transposable elements residing nearby.

### **Identifying the size of the deletion in the *Mdh1-n* or the *Mdh1-n y20* mutants**

Four probes (203G2R, 203G2F, 146B8R, and 45N1R) were hybridized against the *Hind* III digested genomic DNA from two wild-type soybean lines (T322 and Jilin 3) and six mutants with the *Mdh1-n* or the *Mdh1-n y20* allele (T253, T261, T317, T324, T325, and T361). The corresponding BAC clones were used as positive controls. All probes could detect many bands from the *Hind* III-digested soybean genomic DNA (data not shown), which indicated that the sequences of these probes were highly repeated in the soybean genome. Among these probes, 146B8R and 45N1R could not detect clear band patterns. The blots hybridized with these two probes showed smears over the lanes (data not shown). Probe 203G2F could identify a 1.5 Kb *Hind* III band from all the soybean lines that was the same size as the band identified from the positive control, clone 203G2 (data not shown). However, this band was much darker than it was supposed to be if it represented a single-copy gene, which indicated that the band represented multiple genes (data not shown). Thus, the Southern blot results from probes 203G2F, 146B8R, and 45N1R could not determine if the chromosomal regions corresponding to the BAC ends that the probes represented were deleted in the mutants or not.

Probe 203G2R also had many copies in the soybean genome. Southern blot analysis showed that a 2.5 kb band only existed in wild-type soybean lines but not in the mutants (Fig. 5). Since the band was not very sharp, the experiment was repeated several times, and all of them showed the same results. This band corresponded to the reverse end of clone 203G2R, which also was the forward end of the *Mdh1* contig. These results indicated that the fragment corresponding to the reverse end of clone 203G2R was deleted from the mutant genome. Thus, the chromosomal region from the reverse end of clone 203G2R to the *Mdh1* gene with size ~70 kb was deleted from the *Mdh1-n* mutant (T261) and *Mdh1-n y20* mutants (T253, T317, T324, T325, and T361).

However, the break points in mutants could not be identified in this experiment. To identify the break points in mutants, a contig with larger size should be constructed. Alternatively, screening a genome library with a larger insertion such as yeast artificial chromosome (YAC) also would be helpful. A soybean YAC library with an average size of 285 kb and coverage of five soybean genome equivalents was constructed, and by using it, two clones for the *Rps6* region were successfully identified with co-segregating RFLP markers (Santra et al. 2003).

## Conclusions

In the present study, a ~96 kb contig containing the *Mdh1* gene (AF180335) was constructed from three BAC clones (203G2, 146B8, and 45N1) with the reverse end of clone 203G2 and the forward end of clone 146B8 as one end, and the reverse end of clone 45N1 as the other end. Genomic deletions in the *Mdh1-n* mutant (T261) and *Mdh1-n y20* mutants

(T253, T317, T324, T325, and T361) were detected with Southern analysis using the two ends of each clone as probes, except for the forward end of clone 45N1 (45N1F). The results indicated the deletions in the *Mdh1-n* mutant (T261) and *Mdh1-n y20* mutants (T253, T317, T324, T325, and T361) were larger than 70 kb. The break points of the deletion could not be identified. To identify the break point in *Mdh1-n* or *Mdh1-ny20* mutants, the contig for the *Mdh1* gene needs to be longer in length, which could be done by screening BAC libraries with a larger insertion size or screening YAC libraries.

### Acknowledgements

The authors would like to thank Dr. M. K. Battacharyya, Department of Agronomy, Iowa State University, Ames, IA, for providing the Williams 82 BAC library. The authors would also like to thank Dr. C. Gitel for providing probes (watermelon m*Mdh* cDNA) for Southern analysis.

### References

- Amberger LA, Shoemaker RC, and Palmer RG (1992) Inheritance of two independent isozyme variants in soybean plants derived from tissue culture. *Theor Appl Genet* 84: 600-607
- Ashfield T, Bocian A, Held D, Henk AD, Marek LF, Danesh D, Penuela S, Meksem K, Lightfoot DA, Young ND, Shoemaker RC, and Innes RW (2003) Genetic and physical localization of the soybean *Rpg1-b* disease resistance gene reveals a complex locus containing several tightly linked families of NBS-LRR genes. *Mol Plant Microbe Interact* 16: 817-826

- Bennetzen JL (2002) Mechanisms and rates of genome expansion and contraction in flowering plants. *Genetica* 115: 29-36
- Chen XF and Palmer RG (1996) Inheritance and linkage with the *k2* and *Mdh1-n* loci in soybean. *J Hered* 87: 433-347
- Chen XF and Palmer RG (1998a) Recombination and linkage estimation between the *k2* and *Mdh1-n y20* loci in soybean. *J Hered* 89: 488-494
- Chen XF and Palmer RG (1998b) Instability at the *k2 Mdh1-n y20* chromosomal region in soybean. *Mol Gen Genet* 260: 309-318
- Chen XF, Imsande J, and Palmer RG (1999) Eight new mutants at the *k2 Mdh1-n y20* chromosomal region in soybean. *J Hered* 90: 339-403
- Devos KM, Brown JKM, and Bennetzen JL (2002) Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Res* 12: 1075-1079
- Fu H and Dooner HK (2002) Intraspecific violation of genetic colinearity and its implications in maize. *Proc Natl Acad Sci USA* 99: 9573-9578
- Groose RW, Weigelt HD, and Palmer RG (1988) Somatic analysis of an unstable mutation for anthocyanin pigmentation in soybean. *J Hered* 79: 263-267
- Imsande J, Pittig J, Palmer RG, Wimmer C, and Gietl C (2001) Independent spontaneous mitochondrial malate dehydrogenase null mutants in soybean are the result of deletions. *J Hered* 92: 333-338
- Keim P, Olson T, and Shoemaker RC (1988) A rapid protocol for isolating soybean DNA. *Soybean Genet Newsl* 15: 150-152

- Lewers K, Heinz R, Beard H, Marek L, and Matthews B (2002) A physical map of a gene-dense region in soybean linkage group A2 near the black seed coat and Rhg<sub>4</sub> loci. *Theor Appl Genet* 104: 254-260
- Lin JY, Hass-Jacobus B, Sanmiguel P, Walling JG, Yuan Y, Shoemaker RC, Young ND, and Jackson S (2005) Centric regions of soybean (*Glycine max* L. Merr.) chromosomes consist of retroelements and tandemly repeated DNA and are structurally and evolutionarily labile. *Genetics* (In press)
- Ma J, Devos KM, and Bennetzen JL (2004) Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. *Genome Res* 14: 860-869
- Meyers BC, Tingey SV, and Morgante M (2001) Abundance, distribution and transcriptional activity of repetitive elements in the maize genome. *Genome Res* 11: 1660-1676
- Page DR, Kohler C, Da Costa-Nunes JA, Baroux C, and Moore JM (2004) Intrachromosomal excision of a hybrid *Ds* element induces large genomic deletions in *Arabidopsis*. *Proc Natl Acad Sci USA* 101: 2969-2974
- Palmer RG, Burzlaff JD, and Shoemaker RC (2000) Genetic analyses of two independent chlorophyll-deficient mutants identified among the progeny of a single chimeric foliage soybean plant. *J Hered* 91: 297-303
- Palmer RG, Hedges BR, Benavente RS, and Groose RW (1989) *w4*-mutable line in soybean. *Dev Genet* 10: 542-551
- Palmer RG, Pfeiffer TW, Buss GR, and Kilen TC (2004) Qualitative genetics. In Boerma HR, Specht JE (eds) *Soybeans: Improvement, Production, and Uses*, 3rd edn. ASA, CSSA, SSSA, Madison, WI, pp 137-214

- Palmer, RG (1984) Pleiotropy or close linkage of two mutants in soybeans. *J Hered* 75: 457-462
- Sambrook J, Fritsch EF, and Maniatis T (1989) *Molecular cloning: a laboratory manual*. 2nd edn. Cold Spring Harbor Laboratory Press. Plainview, NY
- Santra DK, Sandhu D, Tai T, and Bhattacharyya MK (2003) Construction and characterization of a soybean yeast artificial chromosome library and identification of clones for the *Rps6* region. *Funct Integr Genomics* 3: 153 - 159
- Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y, and Simon M (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc Natl Acad Sci USA* 89: 8794-8797
- Song QJ, Marek LF, Shoemaker RC, Lark KG, Concibido VC, Delannay X, Specht JE, and Cregan PB (2004) A new integrated genetic linkage map of the soybean. *Theor Appl Genet* 109: 122-128
- Vicient CM, Suoniemi A, Ananthawat-Jónsson K, Tanskanen J, Beharav A, Nevo E, and Schulman AH (1999) Retrotransposon *BARE-1* and its role in genome evolution in the genus *Hordeum*. *Plant Cell* 11: 1769-1784
- Woo SS, Jiang J, Gill BS, Paterson AH, and Wing RA (1994) Construction and characterization of a bacterial artificial chromosome library of *Sorghum bicolor*. *Nucleic Acid Res* 22: 4922-4931
- Wright DA and Voytas DF (2002) *Athila4* of *Arabidopsis* and *Calypso* of soybean define a lineage of endogenous plant retroviruses. *Genome Res* 12: 122–131

Table 1. Description of soybean lines used in these experiments

Cultivars	Gene symbol			Discription
	<i>K2</i>	<i>Mdh1</i>	<i>Y20</i>	
Williams	+	+	+	Wild type
T322	+	+	+	Purple and white variegated flowers ( <i>w4-m</i> )
Jilin3 (PI 427.099)	+	+	+	Wild type
T253	<i>k2</i>	<i>Mdh1-n</i>	<i>y20</i>	spontaneous mutation from cv. Harosoy
T261	<i>k2</i>	<i>Mdh1-n</i>	+	spontaneous mutation from cv. Mandarin (Ottawa)
T323 (CD-1)	+	<i>Mdh1-n</i>	<i>y20</i>	genomic revertants from T322.
T324 (CD-2)	+	<i>Mdh1-n</i>	<i>y20</i>	genomic revertants from T322.
T317	+	<i>Mdh1-n</i>	<i>y20</i>	mutated from T317 in tissue culture
T361 (JB-yellow)	+	<i>Mdh1-n</i>	<i>y20</i>	mutated from T317 in tissue culture

Note: *k2*: tan-saddle seed coat, *y20*: yellow foliage, *Mdh1-n*: malate dehydrogenase-1 null

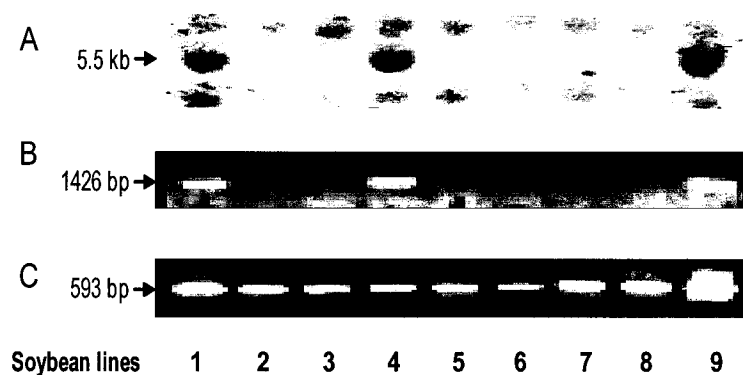


Table 2. PCR primers used to identify specific DNA markers corresponding to the *Mdh1* gene

Marker name	Primers for identifying marker	Position in the <i>Mdh1</i> sequence (AF180335)	Size (bp)
MD42	F: 5'-AGTAGATCCGTATCATGATCCTCCT-3' R: 5'-GTATGGTAGAGTGAAAGTATAATGC-3'	4246 5037	792
MD49	F: 5'-AGGATAAGTAGGAGAACTATGAGC-3' R: 5'-TTTTTCGTCATCTTCCCCATATGGA-3'	4973 5685	713
MD56	F: 5'-TACAATCCATATGGGGAAGATGACG-3' R: 5'-GGATATACATACTGATCGTGTAAGC-3'	5656 6460	805
MD64	F: 5'-TAGAGATTGGCGATAAGCAAAAGGG-3' R: 5'-ATCTGAGCATCGATGGCTTCATCAT-3'	6411 7278	868
Mdh1-5'	F: 5'-ATGATGAAGCCATCGATGCT-3' R: 5'-GCTTTTCCAAGCTCTTCATC-3'	7254 8679	1426
MD85	F: 5'-AAGGTGATGAAGAGCTTGGAAGC-3' R: 5'-TTACACTTGACATACCAGCAACTGG-3'	8555 9350	796
MD92	F: 5'-AAAACCTTCTATGCTGGGAAAGCCA-3' R: 5'-TTGAGGCAAGCATCAGCAAAAAGGG-3'	9296 10097	802
Mdh1-3'	F: 5'-TATGCTGGTGCCCTTTTGC-3' R: 5'-TTTGAGTTCAGGCTTAAGGC-3'	10073 10665	593

Table 3. PCR primers used to amplify the ends of positive BAC clones

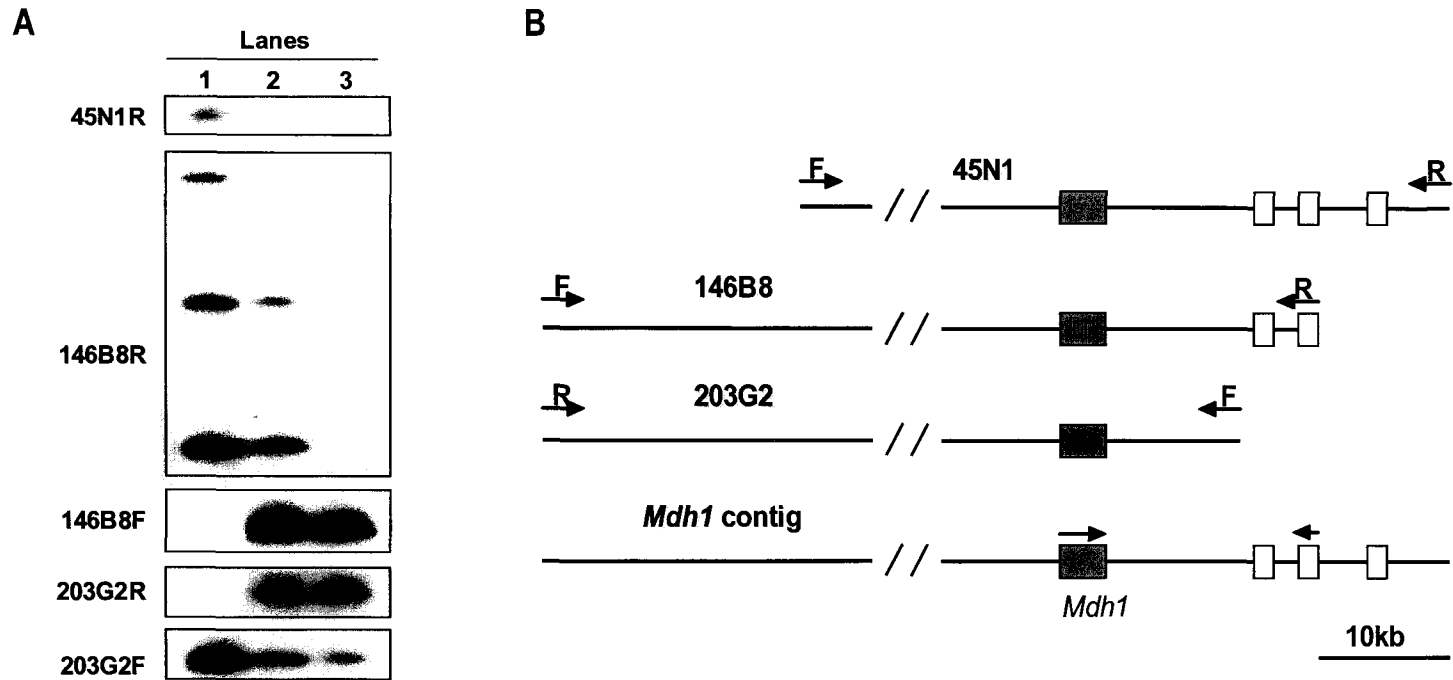
Clone name	Probe name	Primers for amplifying probes	Probe size (bp)
45N1	45N1F	F: 5'-TTTCATTTGAGGATCAAGAG-3' R: 5'-TCTAGTGTATCAAATTTGGA-3'	406
	45N1R	F: 5'-AAATGGAGCTACATTAGTTG-3' R: 5'-TATTAGATGAGTTGAGATCC-3'	383
146B8	146B8F	F: 5'-TGCCTCTCTTTCACCTTGTC-3' R: 5'-ATGATCCAGAAGGTGCTAGG-3'	481
	146B8R	F: 5'-TGTGAGTTGATTTTAGCCTT-3' R: 5'-TAAGTTCCCGAAATCGAACT-3'	428
203G2	203G2F	F: 5'-AACCATAACCGTGAGAGTGT-3' R: 5'-CTTAGATTCTAAGAGAGCAC-3'	510
	203G2R	F: 5'-TAGCTTGCCTCTCTTTCACC-3' R: 5'-TTGTCTCGAGGAGCATAAGG-3'	457



**Fig. 1** Identification of specific markers corresponding to the *Mdh1* gene. Samples from each soybean line presented in figure are: 1. T322 (wild type), 2. T323 (*Mdh1-n y20*), 3. T325 (*Mdh1-n y20*), 4. Jilin3 (wild type), 5. T317 (*Mdh1-n y20*), 6. T361 (*Mdh1-n y20*), 7. T261 (*k2 Mdh1-n*), 8. T253 (*k2 Mdh1-n y20*), 9. Williams (wild type). **A.** Southern analysis of soybean mutants and wild types. Genomic DNA was digested with *EcoRI*, and the probe was watermelon *mMdh* cDNA. A 5.5 kb band detected in wild-type soybean lines is missing in all the examined mutants. **B.** PCR amplification results by using the primers of marker *Mdh1-5'*. The primers could amplify a 1426 bp fragment from wild-type soybean lines but not from the mutants. **C.** PCR amplification results by using the primers of marker *Mdh1-3'*. The primers could amplify a 593bp fragment from all the examined soybean lines.



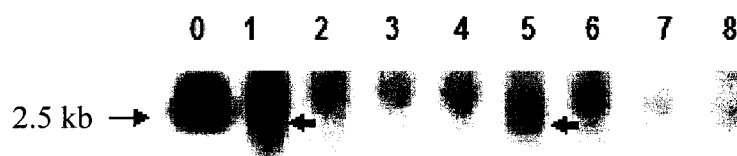
**Fig. 2** Identification of the BAC clone containing the *Mdh1* gene using both PCR and Southern blot analysis. Sixteen BAC clones that gave positive signals using watermelon mMDH as probes were examined. **A.** Screening results with marker *Mdh1*-5'. Three positive clones were identified to contain the *Mdh1* gene. They are 45N1 in lane 2, 146B8 in lane 14, and 203G2 in lane 15. **B.** Southern blot results using watermelon m*Mdh* cDNA as probe. The clones were digested with *EcoRI*. Clones 45N1 in lane 2, 146B8 in lane 14, and 203G2 in lane 15 show a 5.5 kb band corresponding to the *Mdh1* gene. Clones 127L9 and 127L10 show an 8 kb band.



**Fig. 3** Alignment of three BAC clones (45N1, 146B8, and 203G2) containing the *Mdh1* gene. **A.** Southern blot analysis of three clones. Three clones 45N1, 146B8, and 203G2 were digested with *Hind* III, loaded in lanes 1, 2, and 3, respectively, and detected using probes corresponding to the two ends, forward (F) and reverse (R), of each clone. **B.** Alignment of three clones. Three clones were aligned to form a combined *Mdh1* contig with ~96 kb in length. The forward end and reverse end of each clone were referred as F and R, respectively. The *Mdh1* gene was shown as a gray box. *Calypso5-1* like retroelements were presented as a white box. The transcription directions of genes were pointed out with arrows in the combined *Mdh1* contig.

146B8R	1	AGTTGATTTTAGCCTTAGTII TACTTTGGT TATTAGTCAATT-GATC---	46
Calypso5-1	5023	AAITGATTTTAGCCTTAGTII TCACTTTGGT TATTAGTCGATTCGATTAA	5082
146B8R	47	-AAAGGAAACTTCCAAAGAAAAACATTTGATTGATTTTTTGATTATTTT	95
Calypso5-1	5083	GAAAGAGAAATCCCAAGAAAAACGTCGATTTGATTTTTTGATTATTTT	5122
146B8R	96	ATTCAAAGATA-TTTTGATTATTATATTATTTTTCAGATAATTTTGA	144
Calypso5-1	5123	ACTAAAGATATTTTTGATTATTATATTATTTTAC-----	5160
146B8R	145	TTATTTTATTATTATTTTGT-TTTTGGGTTTAAACGAGGTTACAGTGT	193
Calypso5-1	5161	-----CTATTTTGGT TTTCAACG-GGTTACGGCAT	5190
146B8R	194	AAACGATCGGTTAGATTTTGTTTTAAACAAAATTAAAAGGTGTTACAATT	243
Calypso5-1	5191	GACCGAACAGTCGGATTTCATTTTAACAGAAATTAAACGATGTTACAATT	5240
146B8R	244	CAATGATCGGTGGAATTTATTTTATTTT--TTAGGCGAGAAAACGA	291
Calypso5-1	5241	TAAATGATCGGTGGAATTTATTTTATTTT TTAGTTAGGCGAGAAAATGA	5290
146B8R	292	CTTAAATAAATGACTAAAGCACGTCAAAGGGGGTACGGAAGGCCAATGA	341
Calypso5-1	5291	CTTAAATAAATGACTAAAGCACGTCAAAGGGGGTACGGAAGGTAATGA	5340
146B8R	342	AATAAAATAAAAGCACGTGAAACAGTTGGGTCCCTAAGGGTACATAG	391
Calypso5-1	5341	AATGAAAATAAAAGCATGTGAAACAAATGAGGACCTAAGGGTACATA-	5389
146B8R	392	GAATGAATTGAAAAGTTCGATTTTCGGGAAC TTACCGGGTGAAGACCGAAG	441
Calypso5-1	5390	GAATGAATT----GTTTGAITTCGGGAAC TTACCGGTTGAAGATCGAAG	5434
146B8R	442	NACGACGAAGAACGAACNAAGAACGAACNAAGNACTGTGGGGAATCNTT	491
Calypso5-1	5435	AACGACGAAGAACGAACGAAGAACG-TCGATGAACGGT-TGAAAATCTTC	5482
146B8R	492	ACGGAATCCCCCNGGAACATCTCGGAAGC-TTACGGAAGCCCCCGCCT	540
Calypso5-1	5483	GCAAAATCACC-----CACGGAACGTTACGGAAGCACCTCGGCT	5522
146B8R	541	TGGA-TTTCCTTCACGGAACAA-TTTCCTC-CIAATTTT-AGTGAATCTC	586
Calypso5-1	5523	TGGATTTTCTTCACGGAACAAATTTTCTCCTAATTTTAAGTGAATCTC	5572
146B8R	587	AGATACCGAGGAGGATTGAACAAATTG-TCTGCCCT-CTTCCCT--ATTGA	632
Calypso5-1	5573	AGATACCGAGGAGGTCGAACATTTTGTCTTCCCTCTCTTCCCTTATTTA	5622
146B8R	633	TAGGAGAAAAGAGGGAGGTG-GTGCIGCCCCGCTC-CCCCA	670
Calypso5-1	5623	TAGGA-AAAGGAAGGAGATGCTTSCACCCAGCTCGCCCA	5661

**Fig. 4** Alignment of the reverse end of 146B8 (146B8R) and the *G. max* retrovirus-like element *Calypso5-1* (AF186186). The homology between two sequences is 77.8%.



**Fig. 5** Southern analyses of mutants with *Mdh1-n* or *Mdh1-n y20* alleles and their original wild-type parents using probe 203G2R. The samples were loaded in this order: **0.** Clone 203G2 (positive control), **1.** T322 (wild type), **2.** T323 (*Mdh1-n y20*), **3.** T325 (*Mdh1-n y20*), **4.** T253 (*k2 Mdh1-n y20*), **5.** Jilin3 (wild type), **6.** T317 (*Mdh1-n y20*), **7.** T361 (*Mdh1-n y20*) , **8.** T261 (*k2 Mdh1-n*). The 2.5 kb band corresponding to the reverse end of clone 203G2, pointed out with an arrow, was identified in two wild-type soybean lines (T322 and Jilin3), but not in any of the mutants with the *Mdh1* gene deleted.

## CHAPTER 7 GENERAL CONCLUSION

The instability of the *w4-m* locus and the *k2 Mdh1-n y20* chromosomal region were suspected to be due to activities of putative transposable elements. To confirm this hypothesis and to clone the putative transposable elements, several studies were done and presented in this dissertation. They include molecular mapping of these two regions, identifying and genetically analyzing new mutants related to them, identifying gene functions of the *W4* locus, and identifying breakpoints of the deletions in the *Mdh1-n* and *Mdh1-n y20* mutants.

In Chapter 2, a mutation with pale flowers and green hypocotyls, found in self-pollinated progenies of the original mutable line (Asgrow XP2214), was evaluated by crossing to Minsoy (PI 27890), Harosoy *w4* isoline (L72-1138), and Harosoy *w1* isoline (L62-906). The results showed that the pale-flower mutation was conditioned by a new allele at the *W4* locus, namely *w4-p*. The Soybean Genetics Committee assigned Genetic Type Collection number T369 to this true-breeding pale line. Thus, up to now, four alleles have been found at the *W4* locus, and the dominance of these alleles is: *W4* > *w4-m* > *w4-dp*, *w4-p* > *w4*. A F<sub>2</sub> mapping population derived from a single F<sub>1</sub> seed of Minsoy X T369 was used to locate the *W4* locus with SSR markers. The *W4* locus was located on MLG D2 between Satt386 and nearby telomere with 2.3 cM genetic distance apart from Satt386.

In Chapter 3, we tried to determine the function of the *W4* gene in the anthocyanin biosynthesis pathway through biochemical and molecular biological methods. In this study, we checked the accumulation status of flavonol myricetin and anthocyanins in the immature flower petals of six samples from five soybean lines with different *W4* alleles by using HPLC and spectrophotometry, respectively. They are Harosoy (*W4*), Harosoy *w4* (*w4*), T321 (*w4-*



*dp*), T322 (*w4-m*), and T369 (*w4-p*). Here, the white portions of flower petals of T322 (T322w) were collected separately from the purple portion (T322p). The studies indicated that delphinidin and its derivatives were major pigments in soybean flowers. Their contents were reduced in all less-pigmented petals caused by mutations at the *W4* locus. In contrast, the contents of their precursor (dihydromyricetin), reflected by the myricetin contents, increased in these mutant petals. RT-PCR and Northern analysis showed that this phenomenon was related to low transcript level of the DFR2 gene in mutant petals. Further RFLP analysis established that the *W4* gene co-segregated with a DFR gene, which indicated that the *W4* gene would be either a DFR gene or closely linked to one. Based on all the results from this study, the *W4* locus very likely encodes the DFR2 protein.

In Chapter 4, four new mutants were found in the *k2 Mdh1-n y20* region. A new mutation with tan-saddle seed was identified in bulk-harvested seed of cv. Kenwood. SSR analysis showed this mutation was not a contamination from other tan-saddle mutants. Genetic analysis established that this mutation is an allele of the *k2* locus. Three independent mutants with *k2*, *Mdh1-n*, and *y20* alleles were recorded from crossing cv. Williams with T239 (*k2*), which provided more material for further molecular work. But no mutation was found from crosses of Williams with Kenwood-*k2* and T261.

In Chapter 5, we mapped the *k2*, *Mdh1-n*, and *y20* to MLG H with SSR markers using five mapping populations. An integrated map that included all the results from five mapping populations was constructed with software Joinmap 3.0. It consisted of 13 loci in the order Satt541, Satt469, Sat\_122, Satt279, Satt253, Satt314, *Mdh1-n*, *y20*, *k2*, Satt302, Satt142, Satt181, and Satt434. Interestingly, at the position similar to where the *k2 Mdh1-n y20* chromosome region is located, an MDH mobility gene (*Mdh-A/B*) was positioned on the

Iowa State University/USDA map, which is in conflict with the result from the University of Nebraska. On the University of Nebraska map, the same gene was placed on MLG G2. Since the mapping population used by Iowa State University was constructed by *G. soja* X *G. max* instead of a *G. max* intra-species cross, we proposed the the *Mdh-A* genes in *G. soja* and *G. max* were located in MLG H and G2, respectively. In addition, a putative deletion covered three SSR markers (Satt253, Satt279, and Satt314) that was very close to the deletion at the *k2 Mdh1-n* region, was found in T261. And according to the mapping results, these two deletions were separate.

In Chapter 6, we tried to detect the sizes and breakpoints of the deletions in *Mdh1-n* mutant (T261) and *Mdh1-n y20* mutants (T323, T324, T317, T361, and T253). To do this, we constructed a contig that contained the *Mdh1* gene. This contig was composed of three BAC clones (45N1, 146B8, and 203G2) from a Williams 82 BAC library from Dr. Madan K. Battacharyya. Southern analysis was conducted with each end of the clones as a probe against *Hind* III digested genomic DNA from mutants and their original wild-type parents. The result showed that the chromosomal region from the reverse end of clone 203G2 to the *Mdh1* gene with size ~70 kb was deleted in all the mutants. The breakpoints of the deletion could not be determined at this time. In addition, three copies of truncated *Calypso5-1* like retroelements were found downstream of the *Mdh1* gene (AF180335), and clustered in a ~16 kb chromosomal region.

To clarify the mechanism of the instability of the *w4-m* locus and the *k2 Mdh1-n y20* chromosomal region, physically cloning the *w4-m* gene and the breakpoint in the *k2 Mdh1-n y20* chromosomal region are necessary. This study provided more information about these

two regions, generated additional genetic materials, and advanced research towards the final goal.